SPSS 14.0 Base Benutzerhandbuch

Weitere Informationen zu SPSS®-Software-Produkten finden Sie auf unserer Website unter der Adresse *http://www.spss.com*, oder wenden Sie sich an

SPSS Inc.
233 South Wacker Drive, 11th Floor
Chicago, IL 60606-6412, USA
Tel.: (312) 651-3000
Fax: (312) 651-3668

SPSS ist eine eingetragene Marke, und weitere Produktnamen sind Marken der SPSS Inc. für Computerprogramme von SPSS Inc. Die Herstellung oder Verbreitung von Materialien, die diese Programme beschreiben, ist ohne die schriftliche Erlaubnis des Eigentümers der Marke und der Lizenzrechte der Software und der Copyrights der veröffentlichten Materialien verboten.

Die SOFTWARE und die Dokumentation werden mit BESCHRÄNKTEN RECHTEN zur Verfügung gestellt. Verwendung, Vervielfältigung und Veröffentlichung durch die Regierung unterliegt den Beschränkungen in Unterabschnitt (c)(1)(ii) von The Rights in Technical Data and Computer Software unter 52.227-7013. Vertragspartner/Hersteller ist SPSS Inc., 233 South Wacker Drive, 11th Floor, Chicago, IL 60606-6412.

Allgemeiner Hinweis: Andere in diesem Dokument verwendete Produktnamen werden nur zu Identifikationszwecken genannt und können Marken der entsprechenden Unternehmen sein.

TableLook ist eine Marke der SPSS Inc.
Windows ist eine eingetragene Marke der Microsoft Corporation.
DataDirect, DataDirect Connect, INTERSOLV und SequeLink sind eingetragene Marken von DataDirect Technologies.
Teile dieses Produkts wurden unter Verwendung von LEADTOOLS © 1991-2000, LEAD Technologies, Inc., erstellt. ALLE RECHTE VORBEHALTEN.
LEAD, LEADTOOLS und LEADVIEW sind eingetragene Marken von LEAD Technologies, Inc.
Sax Basic ist eine Marke der Sax Software Corporation. Copyright © 1993-2004, Polar Engineering and Consulting. Alle Rechte vorbehalten.
Teile dieses Produkts basieren auf der Arbeit des FreeType-Teams (*http://www.freetype.org*).
Ein Teil der SPSS-Software beinhaltet zlib-Technologie. Copyright © 1995-2002, Jean-loup Gailly und Mark Adler. Die zlib-Software wird in der verfügbaren Form, ohne ausdrückliche oder stillschweigende Gewährleistung bereitgestellt.
Ein Teil der SPSS-Software beinhaltet Sun-Java-Laufzeitbibliotheken. Copyright © 2003, Sun Microsystems, Inc. Alle Rechte vorbehalten. Die Sun-Java-Laufzeitbibliotheken enthalten von RSA Security, Inc., lizenzierten Code. Einige Teile der Bibliotheken wurden von IBM lizenziert und sind unter folgender Adresse erhältlich: *http://oss.software.ibm.com/icu4j/*.

SPSS 14.0 Base Benutzerhandbuch
Copyright © 2005 SPSS Inc.
Alle Rechte vorbehalten.

Ohne schriftliche Erlaubnis der SPSS GmbH Software darf kein Teil dieses Handbuchs für irgendwelche Zwecke oder in irgendeiner Form mit irgendwelchen Mitteln, elektronisch oder mechanisch, mittels Fotokopie, durch Aufzeichnung oder durch andere Informationsspeicherungssysteme reproduziert werden.

1 2 3 4 5 6 7 8 9 0 08 07 06 05

ISBN 1-56827-682-6

Vorwort

SPSS 14.0

SPSS 14.0 ist ein umfassendes System zum Analysieren von Daten. Mit SPSS können Sie Daten aus nahezu allen Dateitypen entnehmen und aus ihnen Berichte in Tabellenform, Diagramme sowie grafische Darstellungen von Verteilungen und Trends, deskriptive Statistiken und komplexe statistische Analysen erstellen.

Dieses Handbuch, *SPSS 14.0 Base Benutzerhandbuch*, dokumentiert die grafische Benutzeroberfläche von SPSS für Windows. Beispiele für statistische Prozeduren in SPSS Base 14.0 finden Sie im Hilfesystem, das mit der Software installiert wird. Algorithmen für die statistischen Prozeduren sind im PDF-Format über das Menü "Hilfe" verfügbar.

Den Menüs und Dialogfeldern von SPSS liegt eine Befehlssprache zugrunde. Auf einige fortgeschrittene Funktionen des System kann nur mit Hilfe der Befehlssyntax zugegriffen werden. (Diese Funktionen sind in der Studentenversion nicht verfügbar.) Detaillierte Informationen zur Befehlssyntax sind auf zwei Arten verfügbar: als Bestandteil der umfassenden Hilfesystems und als separates Dokument im PDF-Format im Handbuch *SPSS 14.0Command Syntax Reference*, das auch über das Menü "Hilfe" verfügbar ist.

SPSS Optionen

Die folgenden Optionen sind als Erweiterungsmodule der Vollversion (nicht der Studentenversion) von SPSS Base erhältlich:

SPSS Regression Models™ bietet Verfahren zur Datenanalyse, die über herkömmliche lineare Statistikmodelle hinausgehen. Es beinhaltet Prozeduren für Probit-Analyse, logistische Regression, Gewichtungsschätzungen, zweistufige Regression kleinster Quadrate und allgemeine nichtlineare Regression.

SPSS Advanced Models™ umfaßt vor allem Verfahren, die in der fortgeschrittenen experimentellen und biomedizinischen Forschung Anwendung finden. Dies beinhaltet beispielsweise Prozeduren für allgemeine lineare Modelle (GLM), lineare gemischte Modelle, Varianz-Komponentenanalyse, loglineare Analysen, ordinale Regression, versicherungsstatistische Sterbetafeln, die Überlebensanalyse nach Kaplan-Meyer sowie die grundlegende und erweiterte Cox-Regression.

SPSS Tables™ dient dem Erstellen einer großen Auswahl von Tabellenberichten in Präsentationsqualität. Mit dieser Option können beispielsweise komplexe Stub- und Banner-Tabellen erstellt und Daten von Mehrfachantworten angezeigt werden.

SPSS Trends™ bietet Funktionen zum Ausführen umfangreicher Prognosen sowie Zeitreihenanalysen mit Modellen für mehrfache Kurvenanpassung, mit Glättungsmodellen und Methoden zum Schätzen autoregressiver Funktionen.

SPSS Categories® bietet Funktionen zum Ausführen und Optimieren von Skalierungsprozeduren, u. a. Korrespondenzanalysen.

SPSS Conjoint™ ermöglicht die Durchführung von Conjoint-Analysen.

SPSS Exact Tests™ berechnet exakte *P*-Werte für statistische Tests bei einer kleinen Anzahl oder sehr ungleichmäßig verteilten Stichproben, bei denen herkömmliche Tests nur ungenaue Ergebnisse liefern.

SPSS Missing Value Analysis™ dient zum Beschreiben von Mustern bei fehlenden Daten, zum Schätzen von Mittelwerten und anderen statistischen Größen sowie zum Ersetzen von Werten für fehlende Beobachtungen.

SPSS Maps™ bereitet geografisch verteilte Daten in Form von hochwertigen Karten mit Symbolen, Farben, Balkendiagrammen, Kreisdiagrammen und Themenkombinationen auf. So können Sie nicht nur das “Was”, sondern auch das “Wo” zeigen.

SPSS Complex Samples™ ermöglicht Experten auf den Gebieten Umfragen, Marktforschung, Gesundheitswesen und Öffentliche Meinung sowie Sozialwissenschaftlern, die das Verfahren der Stichprobenumfrage verwenden, ihre Stichprobenpläne mit komplexen Stichproben in die Datenanalyse zu integrieren.

SPSS Classification Trees™ (Klassifzierungsbäume) Hiermit wird ein baumbasiertes Klassifizierungsmodell erstellt. Die Fälle werden in Gruppen klassifiziert oder es werden Werte für eine abhängige Variable (Zielvariable) auf der Grundlage der Werte von unabhängigen Variablen (Einflussvariablen) vorhergesagt. Die Prozedur umfasst Validierungswerkzeuge für die explorative und die bestätigende Klassifikationsanalyse.

Mit **SPSS Data Validation™** erhalten Sie rasch eine visuelle Ansicht Ihrer Daten. Damit verfügen Sie über die Möglichkeit, Validierungsregeln anzuwenden, mit denen Sie ungültige Datenwerte identifizieren können. Sie können Regeln erstellen, mit denen Werte außerhalb des Bereichs, fehlende Werte oder leere Werte gekennzeichnet werden. Sie können außerdem Variablen speichern, mit denen individuelle Regelverletzungen sowie die Gesamtanzahl von Regelverletzungen je Fall aufgezeichnet werden. Im Lieferumfang des Programms befindet sich ein Satz von vordefinierten Regeln, die Sie kopieren und bearbeiten können.

Zur Produktgruppe von SPSS gehören außerdem Anwendungen für Dateneingabe, Textanalyse, Klassifikation, neurale Netzwerke und das Erstellen von Ablaufdiagrammen.

Installation

Zur Installation von SPSS Base system führen Sie den Lizenzautorisierungsassistenten mit dem Autorisierungscode aus, den Sie von SPSS erhalten haben. Weitere Informationen finden Sie in den Installationsanweisungen im Lieferumfang von SPSS Base system.

Kompatibilität

SPSS kann auf vielen Computersystemen ausgeführt werden. Mindestanforderungen an das System und Empfehlungen finden Sie in den Unterlagen, die mit Ihrem System geliefert werden.

Seriennummern

Die Seriennummer des Programms dient gleichzeitig als Identifikationsnummer bei SPSS. Sie benötigen diese Seriennummer, wenn Sie sich an SPSS wenden, um Informationen über Kundendienst, zu Zahlungen oder Aktualisierungen des Systems zu erhalten. Die Seriennummer wird mit dem Base-System ausgeliefert.

Kundendienst

Wenden Sie sich mit Fragen bezüglich der Lieferung oder Ihres Kundenkontos an Ihr regionales SPSS-Büro. Sie finden die Kontaktdaten auf der SPSS-Website unter *http://www.spss.com/worldwide*. Halten Sie bitte stets Ihre Seriennummer bereit.

Ausbildungsseminare

SPSS bietet öffentliche und unternehmensinterne Seminare an. Alle Seminare beinhalten auch praktische Übungen. Seminare finden in größeren Städten regelmäßig statt. Wenn Sie weitere Informationen zu diesen Schulungen wünschen, wenden Sie sich an Ihr regionales SPSS-Büro, das Sie auf der SPSS-Website unter *http://www.spss.com/worldwide* finden.

Technischer Support

Registrierte Kunden von SPSS können den technischen Support in Anspruch nehmen. Kunden können sich an den technischen Support wenden, wenn sie Hilfe bei der Arbeit mit SPSS oder bei der Installation in einer der unterstützten Hardware-Umgebungen benötigen. Informationen über den technischen Support finden Sie auf der Website von SPSS unter *http://www.spss.com*, oder wenden Sie sich an Ihr regionales SPSS-Büro, das Sie auf der SPSS-Website unter *http://www.spss.com/worldwide* finden. Bei einem Anruf werden Sie nach Ihrem Namen, dem Namen Ihrer Organisation und Ihrer Seriennummer gefragt.

Weitere Veröffentlichungen

Weitere Exemplare von Handbüchern für SPSS-Produkte können direkt bei SPSS Inc. bestellt werden. Besuchen Sie den SPSS Web Store unter *http://www.spss.com/estore*, oder wenden Sie sich an Ihr regionales SPSS-Büro. Sie finden die Kontaktdaten auf der SPSS-Website unter *http://www.spss.com/worldwide*. Wenden Sie sich bei telefonischen Bestellungen in den USA und Kanada unter 800-543-2185 direkt an SPSS Inc. Wenden Sie sich bei telefonischen Bestellungen außerhalb von Nordamerika an Ihr regionales SPSS-Büro, das Sie auf der SPSS-Website finden.

Das Handbuch *SPSS Statistical Procedures Companion* von Marija Norušis wurde von Prentice Hall veröffentlicht. Eine neue Fassung dieses Buchs mit Aktualisierungen für SPSS 14.0 ist geplant. Das Handbuch *SPSS Advanced Statistical Procedures Companion*, bei dem auch SPSS 14.0 berücksichtigt wird, erscheint demnächst. Das Handbuch *SPSS Guide to Data Analysis* für SPSS 14.0 wird ebenfalls derzeit erstellt. Ankündigungen für Veröffentlichungen, die ausschließlich über Prentice Hall verfügbar sind, finden Sie auf der SPSS-Website unter *http://www.spss.com/estore* (wählen Sie Ihr Land aus, und klicken Sie auf Books).

Kundenmeinungen

Ihre Meinung ist uns wichtig. Teilen Sie uns bitte Ihre Erfahrungen mit SPSS-Produkten mit. Insbesondere haben wir Interesse an neuen, interessanten Anwendungsgebieten von SPSS Base system. Senden Sie uns eine E-Mail an *suggest@spss.com*, oder schreiben Sie an: SPSS Inc., Attn: Director of Product Planning, 233 South Wacker Drive, 11th Floor, Chicago, IL 60606-6412.

Über dieses Handbuch

Dieses Handbuch erläutert die grafische Benutzeroberfläche für die Prozeduren in SPSS Base system. Die Abbildungen der Dialogfelder stammen aus SPSS für Windows. Die Dialogfelder unter anderen Betriebssystemen weisen eine ähnliche Gestaltung auf. Detaillierte Informationen zur Befehlssyntax für die Funktionen in SPSS Base system sind auf zwei Arten verfügbar: als Bestandteil des umfassenden Hilfesystems und als separates Dokument im PDF-Format im Handbuch *SPSS 14.0 Command Syntax Reference*, das auch über das Menü “Hilfe” verfügbar ist.

Kontakt zu SPSS

Wenn Sie in unseren Verteiler aufgenommen werden möchten, wenden Sie sich an eines unserer Büros. Sie finden die Kontaktdaten auf der SPSS-Website unter *http://www.spss.com/worldwide*.

Inhalt

1 Übersicht 1

Neuerungen in SPSS 14.0 2
Fenster 7
Menüs 9
Statusleiste 9
Dialogfelder 10
Variablennamen und Variablenlabels in Listen von Dialogfeldern 11
Steuerelemente in Dialogfeldern 11
Untergeordnete Dialogfelder 12
Auswählen von Variablen 12
Symbole für Variablenlisten 13
Aufrufen von Informationen zu Variablen in einem Dialogfeld 13
Grundlegende Schritte bei der Datenanalyse 14
Statistik-Assistent 15
Weitere Informationen zu SPSS 15

2 Aufrufen der Hilfe 17

Verwenden des Inhaltsverzeichnisses der SPSS-Hilfe 20
Verwenden des Hilfe-Indexes 21
Verwenden der Suche in der Hilfe 21
Aufrufen der Hilfe zu Steuerelementen in Dialogfeldern 22
Aufrufen der Hilfe zu ausgegebenen Begriffen 23
Verwenden von Fallstudien 24
Kopieren von Hilfetext aus einem Popup-Fenster 24

3 Datendateien 25

Öffnen einer Datendatei 25
So öffnen Sie Datendateien: 25
Datendateitypen 26
Datei öffnen: Optionen 27
Einlesen von Dateien aus Excel 5 oder nachfolgenden Versionen 27
Einlesen von älteren Excel-Dateien und anderen Tabellenkalkulationsdateien 28
Einlesen von dBASE-Dateien 28
Einlesen von Stata-Dateien 29
Einlesen von Datenbankdateien 29
Text-Assistent 50
Einlesen von Daten aus Dimensions 62
Informationen zur Datei 68
Speichern von Datendateien 69
So speichern Sie geänderte Datendateien: 69
Speichern von Datendateien im Excel-Format 69
Speichern von Datendateien im SAS-Format 70
Speichern von Datendateien im Stata-Format 72
Speichern von Datendateien in anderen Formaten 74
Speichern von Daten: Datendateitypen 74
Speichern von Untergruppen von Variablen 77
Datei speichern: Optionen 78
Schützen der ursprünglichen Daten 78
Virtuelle aktive Datei 78

4 Modus für verteilte Analysen 83

Verteilte Analyse im Vergleich zur lokalen Analyse 84

5 Daten-Editor 95

Datenansicht 96
Variablenansicht. 97
Eingeben von Daten 110
Bearbeiten von Daten 113
Gehe zu Fall 116
Status für die Fallauswahl im Daten-Editor 117
Optionen für die Anzeige im Daten-Editor 118
Drucken aus dem Daten-Editor 120

6 Arbeiten mit mehreren Datenquellen 121

Grundsätzlicher Umgang mit mehreren Datenquellen 122
Kopieren und Einfügen von Informationen zwischen Daten-Sets 124
Umbenennen von Daten-Sets 124

7 Aufbereitung von Daten 125

Variableneigenschaften 125
Definieren von Variableneigenschaften 126
Mehrfachantworten-Sets 134
Kopieren von Dateneigenschaften 138

Ermitteln doppelter Fälle 149
Bereichseinteiler 152

8 Transformieren von Daten 165

Berechnen von Variablen 165
Funktionen 169
Fehlende Werte in Funktionen. 170
Zufallszahlengeneratoren 170
Häufigkeiten von Werten in Fällen zählen 171
Umkodieren von Werten 174
Umkodieren in dieselben Variablen 174
Umkodieren in andere Variablen 177
Rangfolge bilden. 181
Automatisch umkodieren 184
Assistent für Datum und Uhrzeit 188
Datentransformationen für Zeitreihen 206
Bewerten von Daten mit Vorhersagemodellen 214

9 Umgang mit Dateien und Dateitransformationen 221

Fälle sortieren. 222
Transponieren 223
Zusammenfügen von Datendateien. 223
Fälle hinzufügen 224
Variablen hinzufügen 229
Daten aggregieren 232

Datei aufteilen 237
Fälle auswählen 238
Fälle gewichten 243
Umstrukturieren von Daten 245

10 Arbeiten mit Ausgaben 271

Viewer 271
Verwenden von Ausgaben in anderen Anwendungen 280
Einfügen von Objekten im Viewer 284
Inhalte einfügen 284
Einfügen von Objekten aus anderen Anwendungen in den Viewer. 285
Ausgabe exportieren 285
Ausdrucken von Viewer-Dokumenten 297
Speichern der Ausgabe 305

11 Text-Viewer 307

So erstellen Sie eine Textausgabe: 308
Steuern des Textausgabeformats 309
Eigenschaften der Schriftart 314
So drucken Sie eine Textausgabe 314
So speichern Sie Text-Viewer-Ausgaben 316

12 Pivot-Tabellen 317

Bearbeiten von Pivot-Tabellen 317

Arbeiten mit Schichten 322
Lesezeichen 326
Ein- und Ausblenden von Zellen 327
Bearbeiten von Ergebnissen 329
Ändern der Darstellung von Tabellen 330
Tabelleneigenschaften 332
So ändern Sie die Eigenschaften von Pivot-Tabellen 332
Tabelleneigenschaften: Allgemein 333
Tabelleneigenschaften: Fußnoten 334
Tabelleneigenschaften: Zellenformate 335
Tabelleneigenschaften: Rahmen 337
Tabelleneigenschaften: Drucken 338
Schriftart 339
Breite der Datenzellen 340
Zelleneigenschaften 342
Zelleneigenschaften: Wert 342
Zelleneigenschaften: Ausrichtung 344
Zelleneigenschaften: Ränder 345
Zelleneigenschaften: Schattierung 346
Fußnotenzeichen 346
Auswählen von Zeilen und Spalten in Pivot-Tabellen 347
So wählen Sie eine Zeile oder Spalte in einer Pivot-Tabelle aus 348
Anpassen von Ergebnissen in Pivot-Tabellen 348
Drucken von Pivot-Tabellen 349
So drucken Sie die verborgenen Schichten von Pivot-Tabellen 350
Festlegen von Tabellenumbrüchen für breite und lange Tabellen 350

13 Arbeiten mit der Befehlssyntax 353

Regeln für die Syntax 353
Übernehmen der Befehlssyntax aus Dialogfeldern 356
Kopieren von Syntax aus dem Ausgabe-Log 357
Bearbeiten der Syntax in einer Journaldatei 359
So führen Sie Befehlssyntax aus. 361
Mehrere Execute-Befehle 362

14 Häufigkeiten 363

Häufigkeiten: Statistik. 366
Häufigkeiten: Diagramme 368
Häufigkeiten: Format 369

15 Deskriptive Statistiken 371

Deskriptive Statistik: Optionen 373
Zusätzliche Funktionen beim Befehl DESCRIPTIVES 374

16 Explorative Datenanalyse 377

Explorative Datenanalyse: Statistik 381
Explorative Datenanalyse: Diagramme 382
Explorative Datenanalyse: Optionen 384
Zusätzliche Funktionen beim Befehl EXAMINE 384

17 Kreuztabellen 387

Kreuztabellenschichten 390
Kreuztabellen: Gruppierte Balkendiagramme 391
Kreuztabellen: Statistik 391
Kreuztabellen: Zellenanzeige 395
Kreuztabellen: Tabellenformat 397

18 Zusammenfassen 399

Zusammenfassen: Optionen 402
Zusammenfassung: Statistik 403

19 Mittelwerte 407

Mittelwerte: Optionen 410

20 OLAP-Würfel 413

OLAP-Würfel: Statistiken 416
OLAP-Würfel: Differenzen 419
OLAP-Würfel: Titel 420

21 T-Tests 421

T-Test bei unabhängigen Stichproben 421

T-Test bei gepaarten Stichproben 425
T-Test bei einer Stichprobe 428
Zusätzliche Funktionen beim Befehl T-TEST 431

22 Einfaktorielle ANOVA 433

Einfaktorielle ANOVA: Kontraste 436
Einfaktorielle ANOVA: Post-Hoc-Mehrfachvergleiche 437
Einfaktorielle ANOVA: Optionen 440
Zusätzliche Funktionen beim Befehl ONEWAY 441

23 GLM - Univariat 443

GLM: Modell 447
GLM: Kontraste 450
GLM: Profilplots 451
GLM: Post-Hoc-Vergleiche 453
GLM: Speichern 456
GLM-Optionen 458
Zusätzliche Funktionen beim Befehl UNIANOVA 460

24 Bivariate Korrelationen 463

Bivariate Korrelationen: Optionen 466
Zusätzliche Funktionen bei den Befehlen CORRELATIONS und NONPAR CORR 467

25 Partielle Korrelationen 469

Partielle Korrelationen: Optionen 472
Zusätzliche Funktionen beim Befehl PARTIAL CORR 473

26 Distanzen 475

Unähnlichkeitsmaße für Distanzen 477
Ähnlichkeitsmaße für Distanzen 478
Zusätzliche Funktionen beim Befehl PROXIMITIES 479

27 Lineare Regression 481

Lineare Regression: Methode zur Auswahl von Variablen 485
Lineare Regression: Bedingung aufstellen 487
Lineare Regression: Diagramme 487
Lineare Regression: Speichern von neuen Variablen 489
Lineare Regression: Statistiken 493
Lineare Regression: Optionen 495
Zusätzliche Funktionen beim Befehl REGRESSION 496

28 Kurvenanpassung 497

Modelle für die Kurvenanpassung 501
Kurvenanpassung: Speichern 502

29 Diskriminanzanalyse 505

Diskriminanzanalyse: Bereich definieren 508
Diskriminanzanalyse: Fälle auswählen 508
Diskriminanzanalyse: Statistik 509
Diskriminanzanalyse: Schrittweise Methode. 511
Diskirminanzanalyse: Klassifizieren 512
Diskriminanzanalyse: Speichern 514
Zusätzliche Funktionen beim Befehl DISCRIMINANT 515

30 Faktorenanalyse 517

Faktorenanalyse: Fälle auswählen 523
Faktorenanalyse: Deskriptive Statistiken. 523
Faktorenanalyse: Extraktion 524
Faktorenanalyse: Rotation 527
Faktorenanalyse: Faktorwerte 528
Faktorenanalyse: Optionen 529
Zusätzliche Funktionen beim Befehl FACTOR. 530

31 Auswählen einer Prozedur zum Durchführen einer Clusteranalyse 531

32 Two-Step-Clusteranalyse 533

Two-Step-Clusteranalyse: Optionen 537

Two-Step-Clusteranalyse: Diagramme 540
Two-Step-Clusteranalyse: Ausgabe 541

33 Hierarchische Clusteranalyse 543

Hierarchische Clusteranalyse: Methode 547
Hierarchische Clusteranalyse: Statistik 548
Hierarchische Clusteranalyse: Diagramme 549
Hierarchische Clusteranalyse: Neue Variablen 549
Zusätzliche Funktionen beim Befehl CLUSTER 550

34 Clusterzentrenanalyse 551

Clusterzentrenanalyse: Effizienz 556
Clusterzentrenanalyse: Iterieren 557
Clusterzentrenanalyse: Neue Variablen 558
Clusterzentrenanalyse: Optionen 558
Zusätzliche Funktionen beim Befehl QUICK CLUSTER 559

35 Nichtparametrische Tests 561

Chi-Quadrat-Test 562
Test auf Binomialverteilung 566
Sequenzentest 569
Kolmogorov-Smirnov-Test bei einer Stichprobe 573
Tests bei zwei unabhängigen Stichproben 576
Tests bei zwei verbundenen Stichproben 580

Tests bei mehreren unabhängigen Stichproben 584
Tests bei mehreren verbundenen Stichproben 588

36 Analyse von Mehrfachantworten 593

Mehrfachantworten: Sets definieren 594
Mehrfachantworten: Häufigkeiten 596
Mehrfachantworten: Kreuztabellen 598
Mehrfachantworten: Kreuztabellen, Bereich definieren 601
Mehrfachantworten: Kreuztabellen, Optionen. 601
Zusätzliche Funktionen beim Befehl MULT RESPONSE 603

37 Ergebnisberichte 605

Bericht in Zeilen 605
Bericht in Spalten. 613
Zusätzliche Funktionen beim Befehl REPORT 619

38 Reliabilitätsanalyse 621

Reliabilitätsanalyse: Statistik 623
Zusätzliche Funktionen beim Befehl RELIABILITY 626

39 Multidimensionale Skalierung 627

Multidimensionale Skalierung: Form der Daten. 630

Multidimensionale Skalierung: Distanzen aus Daten erstellen. 630
Multidimensionale Skalierung: Modell 631
Multidimensionale Skalierung: Optionen. 633
Zusätzliche Funktionen beim Befehl ALSCAL. 633

40 Verhältnisstatistik 635

Verhältnisstatistik 637

41 Übersicht über die Diagrammfunktion 639

Erstellen und Ändern von Diagrammen 639
Optionen für die Diagrammdefinition. 646

42 ROC-Kurven 653

ROC-Kurve: Optionen 656

43 Extras 657

Variablenbeschreibungen. 657
Datendateikommentare 658
Variablen-Sets 659
Variablen-Sets definieren 659
Sets verwenden 661
Umsortieren von Listen mit Zielvariablen. 662

44 Optionen 665

Optionen: Allgemein 666
Optionen: Viewer 668
Optionen: Text-Viewer 669
Optionen: Beschriftung der Ausgabe 671
Diagrammoptionen 673
Optionen: Interaktiv 678
Pivottabellenoptionen 679
Optionen: Daten 681
Optionen: Währung 683
Optionen: Skripts 684

45 Anpassen von Menüs und Symbolleisten 687

Menü-Editor 687
Anpassen von Symbolleisten 688
Symbolleisten anzeigen 689
So passen Sie Symbolleisten an:. 689

46 SPSS-Produktionsmodus 695

Verwenden des Produktionsmodus 697
Export-Optionen 698
Benutzerdefinierte Eingabeaufforderungen. 700
Makro-Eingabeaufforderungen im Produktionsmodus. 702
Produktionsmodus: Optionen 703
Festlegen von Formaten für Produktionsjobs. 704

Ausführen von Produktionsjobs aus der Befehlszeile 707
Im Web veröffentlichen 709
SmartViewer-Web-Server: Anmeldung 710

47 Skriptfunktionen in SPSS 711

So führen Sie Skripts aus 711
Im Lieferumfang von SPSS enthaltene Skripts 712
Autoskripts 713
Erstellen und Bearbeiten von Skripts 715
So bearbeiten Sie ein Skript 715
Skript-Fenster 716
Starter-Skripts 718
Erstellen von Autoskripts 720
Funktionsweise von Skripts 725
SPSS-Objektklassen und Konventionen für die Benennung von Variablen . 727
Neue Prozeduren (Skripts) 732
Hinzufügen einer Beschreibung zu einem Skript 735
Skripts mit benutzerdefinierten Dialogfeldern 735
Fehlersuche in Skripts 740
Skriptdateien und Syntaxdateien 743

48 Ausgabeverwaltungssystem (OMS) 747

Ausgabeobjekttypen 751
Befehls-IDs und Tabellenuntertypen 753
Labels 755
OMS: Options 757

Protokollierung 762
Ausschließen der Ausgabeanzeige aus dem Viewer 763
Weiterleiten der Ausgabe an SPSS-Datendateien 763
OXML-Tabellenstruktur 774
OMS-IDs 778

Anhänge

A Administrator für Datenbankzugriff 781

B Anpassen von HTML-Dokumenten 783

So fügen Sie benutzerdefinierten HTML-Code in exportierte Viewer-Dokumente ein:. 783
Inhalt und Format einer Textdatei für benutzerdefinierten HTML-Code . . . 784
So verwenden Sie eine andere Datei oder einen anderen Speicherort für den benutzerdefinierten HTML-Code: 784

Index 787

Kapitel
1

Übersicht

Mit SPSS für Windows verfügen Sie über ein leistungsfähiges System für statistische Analyse und Datenmanagement mit einer grafischen Benutzeroberfläche. Aussagekräftige Menüs und übersichtlich gestaltete Dialogfelder nehmen Ihnen einen großen Teil Ihrer Arbeit ab. Die meisten Aufgaben können einfach mit der Maus durchgeführt werden.

Neben der einfach zu bedienenden Benutzeroberfläche für die statistische Analyse finden Sie in SPSS für Windows die folgenden Hilfsmittel:

Daten-Editor. Der Daten-Editor ist ein vielseitiges System (ähnlich einer Tabellenkalkulation) für das Definieren, Eingeben, Bearbeiten und Anzeigen von Daten.

Viewer. Der Viewer erleichtert das Betrachten der Ergebnisse, das Ein- bzw. Ausblenden der Ausgaben, das Ändern der Anzeigereihenfolge der Ergebnisse und die Übertragung von Tabellen und Diagrammen in Präsentationsqualität von SPSS zu anderen Anwendungen.

Mehrdimensionale Pivot-Tabellen. Mit mehrdimensionalen Pivot-Tabellen erwecken Sie Ihre Ergebnisse zum Leben. Sie können die Anordnung der Zeilen, Spalten und Schichten zum Auswerten Ihrer Tabellen ändern. So können Sie wichtige Ergebnisse hervorheben, die in "normalen" Berichten untergehen würden. Wenn Sie die Tabelle aufteilen, so daß immer nur eine Gruppe angezeigt wird, können Sie Gruppen leichter vergleichen.

Hochauflösende Grafiken. Hochauflösende und farbige Kreisdiagramme, Balkendiagramme, Histogramme, Streudiagramme, 3D-Grafiken und mehr sind jetzt als Standardfunktionen in SPSS enthalten.

Datenbankzugriff. Sie können Informationen aus Datenbanken abrufen, indem Sie anstelle von komplizierten SQL-Abfragen den Datenbank-Assistenten verwenden.

Transformieren von Daten. Die Funktionen für das Transformieren von Daten erleichtern Ihnen die Vorbereitung Ihrer Daten für die Analyse. Unter anderem können Sie Daten problemlos in Untergruppen aufteilen und Kategorien kombinieren sowie Dateien hinzufügen, aggregieren, zusammenfügen, aufteilen und transponieren.

Verteilung auf elektronischem Wege. Mit einem Mausklick können Sie Berichte per E-Mail versenden und Tabellen oder Diagramme im HTML-Format im Internet oder einem Intranet verfügbar machen.

Online-Hilfe. Die ausführlichen Lernprogramme bieten Ihnen einen umfassenden Überblick über das Programm, kontextsensitive Hilfethemen in Dialogfeldern führen Sie durch bestimmte Aufgaben, und Popup-Definitionen in den Ergebnissen von Pivot-Tabellen erklären statistische Begriffe. Der Statistik-Assistent hilft Ihnen, geeignete Prozeduren zu finden, und die Fallstudien enthalten praktische Beispiele zum Verwenden von statistischen Prozeduren und Interpretieren der Ergebnisse.

Befehlssprache. Die meisten Aufgaben können einfach mit der Maus durchgeführt werden. Dennoch bietet SPSS auch eine leistungsfähige Befehlssprache, mit deren Hilfe Sie viele häufig durchzuführende Aufgaben speichern und automatisieren können. Die Befehlssprache bietet außerdem einige Funktionen, die nicht über die Menüs und Dialogfelder zur Verfügung stehen.

Eine vollständige Dokumentation zur Befehlssyntax findet sich im umfassenden Hilfesystem und als separates Dokument im PDF-Format in der *SPSS Command Syntax Reference*, die auch über das Menü “Hilfe” verfügbar ist.

Neuerungen in SPSS 14.0

Datenverwaltung

- Sie können gleichzeitig mehrere Datenquellen öffnen. Dies erleichtert das Vergleichen von Datendateien, das Kopieren von Daten und Attributen von einer Datei in eine andere und das Zusammenfügen von mehreren Datenquellen, ohne zuerst jede Datenquelle als sortierte SPSS-Datendatei zu speichern.
- Datendateien im Stata-Format können eingelesen und geschrieben werden. Sie können Datendateien im Format von Stata, Versionen 4-8, einlesen und Datendateien im Format von Stata, Versionen 5-8, schreiben. Weitere Informationen erhalten Sie, wenn Sie im Hilfessystem auf der Registerkarte “Index” den Begriff Stata eingeben.

- Sie können Daten aus SPSS Dimensions-Datenquellen einlesen, einschließlich Quanvert, Quancept und mrInterview. Für weitere Informationen siehe “Einlesen von Daten aus Dimensions” in Kapitel 3 auf S. 62.
- Sie können Daten aus OLE DB-Datenquellen einlesen. Für weitere Informationen siehe “Auswählen einer Datenquelle” in Kapitel 3 auf S. 31.
- Sie können aussagekräftige Wertelabels definieren, die bis zu 120 Byte lang sind (frühere Obergrenze: 60 Byte).
- Sie können Datenwerte aus Wertelabels erstellen oder Wertelabels mit der Funktion `VALUELABEL` in der Transformationslogik verwenden.
- Mit der Funktion `REPLACE` können Sie String-Werte suchen und ersetzen.
- Sie können benutzerdefinierte Variablen-Attribute und Datendatei-Attribute mit den Befehlen `VARIABLE ATTRIBUTE` und `DATAFILE ATTRIBUTE` definieren.
- Daten können unter Verwendung von Feld- bzw. Spaltennamen, die nicht von den SPSS-Regeln für die Bezeichnung von Variablen eingeschränkt werden, in Datenbanktabellen und Dateien anderer Formate geschrieben werden. `SAVE TRANSLATE` wurde erweitert, sodass Werte in Anführungszeichen für Feld- bzw. Spaltennamen verwendet werden können, die Leerzeichen, Kommata und andere Zeichen enthalten, die in SPSS-Variablennamen nicht zulässig sind.
- Mit dem neuen Unterbefehl `SQL` des Befehls `SAVE TRANSLATE` können neue Spalten an Datenbanktabellen angehängt, Spaltenattribute von Datenbanktabellen geändert, Tabellen verknüpft und andere Vorgänge durchgeführt werden, die in gültigen SQL-Anweisungen zulässig sind.

Diagramme

- In der neuen Benutzeroberfläche der Diagrammerstellung (Menü “Grafiken”) können Sie aus vordefinierten Galeriediagrammen oder aus einzelnen Bestandteilen (z. B. Koordinatensysteme und Balken) Diagramme aufbauen.
- Mit den leistungsfähigen Befehlen `GGRAPH` und `GPL` können Sie benutzerdefinierte Diagrammtypen erstellen.

Erweiterungen bei Statistiken

- Die Option “Trends” enthält den neuen Expert Modeler, der automatisch das Modell der besten Anpassung für eine oder mehrere Zeitreihen ermittelt und schätzt. So müssen Sie sich nicht mehr an das geeignete Modell herantasten. Der

Expert Modeler ist über das Dialogfeld “Zeitreihenmodellierung” oder über die Befehlssyntax (Befehl TSMODEL) verfügbar.

- Mit der neuen Option “Data Validation” erhalten Sie rasch eine visuelle Ansicht Ihrer Daten, und Sie können Validierungsregeln zum Identifizieren von ungültigen Datenwerten anwenden. Sie können Regeln erstellen, mit denen Werte außerhalb des Bereichs, fehlende Werte oder leere Werte gekennzeichnet werden. Sie können außerdem Variablen speichern, mit denen individuelle Regelverletzungen sowie die Gesamtanzahl von Regelverletzungen je Fall aufgezeichnet werden. Im Lieferumfang des Programms befindet sich ein Satz von vordefinierten Regeln, die Sie kopieren und bearbeiten können. Die Datenvalidierung ist über das Dialogfeld “Daten validieren” im Menü “Daten” oder über die Befehlssyntax (Befehl VALIDATEDATA) verfügbar.

  Mit der neuen Prozedur “Anomalie-Erkennung” in der Option “Data Validation” können Sie ungewöhnliche Beobachtungen identifizieren, die Vorhersagemodelle nachteilig beeinflussen könnten. Einige dieser Randbeobachtungen stellen wirklich einzigartige Fälle dar und eignen sich deswegen nicht für eine Vorhersage. Andere Beobachtungen stellen Dateneingabefehler dar, wobei die Werte technisch gesehen “richtig” sind und deswegen nicht mit der Prozedur “Daten validieren” abgefangen werden können. Die Anomalie-Erkennung ist über das Dialogfeld “Ungewöhnliche Fälle identifizieren” im Menü “Daten” oder über die Befehlssyntax (Befehl DETECTANOMALY) verfügbar.

- Bei der neuen Prozedur “Multidimensionale Entfaltung” (PREFSCAL) in der Option “Categories” wird versucht, die Struktur in einem Set von Ähnlichkeitsmaßen zwischen Zeilen- und Spaltenobjekten zu ermitteln. Dies wird durch das Zuweisen von Beobachtungen zu bestimmten Positionen in einem konzeptuellen Raum mit wenigen Dimensionen erzielt, wobei die Abstände zwischen den Punkten des Raumes mit den vorgegebenen Unähnlichkeiten bzw. Ähnlichkeiten so gut wie möglich übereinstimmen. Als Ergebnis werden die Objekte in diesem schwachdimensionierten Raum mithilfe der Methode der kleinsten Quadrate dargestellt, was häufig zu einem besseren Verständnis der Daten beiträgt. Diese Prozedur steht gegenwärtig über den Befehl PREFSCAL der Befehlssyntax zur Verfügung.
- Mit der neuen Prozedur “Prädiktorauswahl” (SELECTPRED) in SPSS Server kann eine äußerst große Anzahl an kategorialen und stetigen Prädiktorvariablen durchsucht werden. Die Prozedur wählt eine kleinere Teilmenge für die Verwendung in Prozeduren für Vorhersagemodelle aus, bei denen nicht so viele

Prädiktoren verwendet werden können. Diese Prozedur steht gegenwärtig über den Befehl SELECTPRED der Befehlssyntax zur Verfügung.

- Die neue Prozedur "Naive Bayes-Klassifikation" (NAIVEBAYES) in SPSS Server erstellt ein einfaches und robustes Modell für die Prädiktorauswahl und Klassifikation. Diese Prozedur steht gegenwärtig über den Befehl NAIVEBAYES der Befehlssyntax zur Verfügung.
- Mit den erweiterten Funktionen für Signifikanztests in der Option "Tables" können Sie jetzt Signifikanztests für Zwischenergebnisse und Mehrfachantworten-Tests durchführen.
- Beim Definieren von Mehrfachantworten-Sets für mehrere Dichotomien haben Sie jetzt mehr Flexibilität.

Ausgabe

- Die Ausgabe für die Prozeduren "Rangfolge bilden" (RANK), "Fehlende Werte ersetzen" (RMV), und "Zeitreihe erstellen" (CREATE) im Base-System, alle Prozeduren in der Option "Conjoint", die Prozedur "Modell für loglineare Analyse auswählen" (HILOGLINEAR) in der Option "Advanced Models" und die Prozeduren "Probit-Analyse", (PROBIT), "Gewichtungsschätzung" (WLS) und "Zweistufige kleinste Quadrate" (2SLS) in der Option "Regression Models" erfolgt jetzt als Pivot-Tabelle.

Leistungsverbesserungen.

- Tabellenstrukturen in der Option "Custom Tables" (CTABLES), bei denen zuvor ein längerer Zeitraum für die Erstellung benötigt oder die Speicherkapazität überschritten wurde, werden jetzt schnell und effizient erstellt.

Verbessertes Erscheinungsbild

- Verbesserte Variablensymbole stellen jetzt auf einen Blick mehr Informationen über Variablen bereit, u. a. das Messniveau (nominal, ordinal, metrisch) und den Datentyp (String, numerisch, Datum, Zeit).
- Windows XP und Themen werden jetzt vollständig unterstützt.

Kompatibilität von SPSS 14.0 mit früheren Versionen

Funktionen ANY und RANGE

In früheren Versionen gaben die Funktionen ANY und RANGE nur einen fehlenden Wert aus, wenn das erste Argument einem fehlenden Wert entsprach. Zur Vereinheitlichung mit der Funktionsweise anderer Funktionen und Berechnungen geben diese Funktionen auch dann einen fehlenden Wert zurück, wenn eines der verbleibenden Argumente systemdefiniert fehlend oder benutzerdefiniert fehlend ist und das erste Argument nicht zu einem von den anderen nichtfehlenden Argumenten passt. Daher entspricht

```
COMPUTE newvar=ANY(var1, var2, var3)
```

jetzt:

```
COMPUTE newvar=(var1=var2 or var1=var3).
```

Logistische Regression

In früheren Versionen von SPSS hing die Reihenfolge von neu kodierten String-Werten von der Reihenfolge der Werte in der Datendatei ab. Wenn zum Beispiel die abhängige Variable neu kodiert wurde, wurde der erste angetroffene String-Wert mit 0 und der zweite angetroffene String-Wert mit 1 neu kodiert. Die Prozedur kodiert String-Variablen jetzt dergestalt, dass die Reihenfolge der neu kodierten Werte der alphanumerischen Reihenfolge der String-Werte entspricht. Deswegen kann es vorkommen, dass die Prozedur String-Variablen anders als in früheren Versionen kodiert. Die logistische Regression ist in der Option “Regression Models” verfügbar.

Makrofunktionen

Verbesserung an den Makrofunktionen können zu Fehlern bei Makros führen, die zuvor fehlerfrei ausgeführt wurden. Wenn ein Makro am Ende eines Befehls aufgerufen wird und sich dort kein Befehlsabschluss befindet (entweder eine Punkt oder eine leere Zeile), wird insbesondere bei Syntax mit interaktiven Regeln der nächste Befehl nach der Makro-Erweiterung als fortgesetzte Zeile anstelle eines neuen Befehls interpretiert, wie in:

```
DEFINE !macro1()
var1 var2 var3
!ENDDEFINE.
FREQUENCIES VARIABLES = !macro1
DESCRIPTIVES VARIABLES = !macro1.
```

Im interaktiven Modus wird der Befehl `DESCRIPTIVES` als fortgesetzte Zeile des Befehls `FREQUENCIES` interpretiert, und keiner der beiden Befehle wird ausgeführt.

Fenster

In SPSS gibt es verschiedene Arten von Fenstern:

Daten-Editor. Der Daten-Editor zeigt den Inhalt der Datendatei an. Im Daten-Editor können Sie neue Datendateien erstellen und vorhandene Datendateien bearbeiten. Wenn Sie mehr als eine Datendatei geöffnet haben, besitzt jede Datendatei ein separates Fenster im Daten-Editor.

Viewer. Alle statistischen Ergebnisse, Tabellen und Diagramme werden im Viewer angezeigt. Sie können die Ausgaben bearbeiten und zur späteren Verwendung speichern. Das Fenster des Viewers wird automatisch geöffnet, wenn Sie das erste Mal eine Prozedur aufrufen, die eine Ausgabe erzeugt.

Text-Viewer. Im Text-Viewer können Sie die Ausgabe auch als einfachen Text (anstelle von interaktiven Pivot-Tabellen) anzeigen lassen.

Pivot-Tabellen-Editor. Im Pivot-Tabellen-Editor verfügen Sie über vielseitige Möglichkeiten zur Bearbeitung von Ausgaben, die als Pivot-Tabellen angezeigt werden. Sie können Text bearbeiten, die Daten in Zeilen und Spalten austauschen, Farben hinzufügen, mehrdimensionale Tabellen erstellen und Ergebnisse ein- bzw. ausblenden.

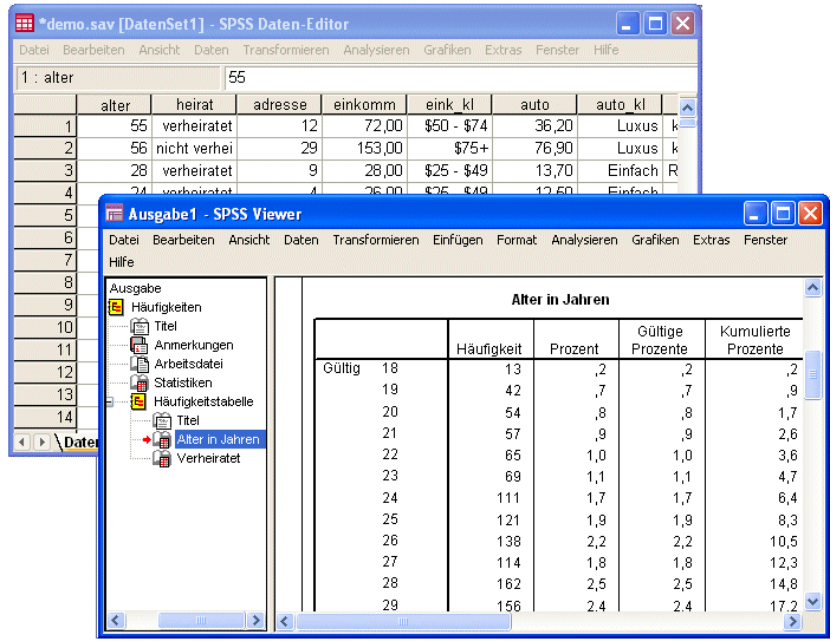
Diagramm-Editor. In Diagrammfenstern können hochauflösende Diagramme und Grafiken bearbeitet werden. Sie können die Farben ändern, andere Schriftarten oder -größen auswählen, horizontale und vertikale Achsen vertauschen, 3D-Streudiagramme rotieren und den Diagrammtyp ändern.

Textausgabe-Editor. Textausgaben, die nicht in Pivot-Tabellen angezeigt werden, können im Textausgabe-Editor bearbeitet werden. Sie können die Ausgabe bearbeiten und die Schriftmerkmale ändern (Schriftart, Stil, Farbe, Größe).

Syntax-Editor. Sie können die in einem Dialogfeld getroffene Auswahl auch als Befehlssyntax direkt in ein Syntax-Fenster einfügen. Dort werden die Befehle als Befehlssyntax angezeigt. Sie können die Befehlssyntax dann bearbeiten und so die Funktionen von SPSS nutzen, auf die Sie nicht über Dialogfelder zugreifen können. Diese Befehle können zur späteren Nutzung in einer Datei gespeichert werden.

Skript-Editor. Mit Skripts und OLE-Automatisierung können Sie viele Aufgaben in SPSS an Ihre Bedürfnisse anpassen und automatisieren. Verwenden Sie den Skript-Editor zum Erstellen und Bearbeiten von Basic-Skripts.

Abbildung 1-1
Daten-Editor und Viewer


*demo.sav [DatenSet1] - SPSS Daten-Editor
Datei Bearbeiten Ansicht Daten Transformieren Analysieren Grafiken Extras Fenster Hilfe
1 : alter
55
alter heirat adresse einkomm eink_kl auto auto_kl
1 55 verheiratet 12 72,00 $50 - $74 36,20 Luxus
2 56 nicht verhei 29 153,00 $75+ 76,90 Luxus
3 28 verheiratet 9 28,00 $25 - $49 13,70 Einfach
Ausgabe1 - SPSS Viewer
Datei Bearbeiten Ansicht Daten Transformieren Einfügen Format Analysieren Grafiken Extras Fenster Hilfe
Ausgabe
Häufigkeiten
Titel
Anmerkungen
Arbeitsdatei
Statistiken
Häufigkeitstabelle
Titel
Alter in Jahren
Verheiratet
Alter in Jahren
Häufigkeit Prozent Gültige Prozente Kumulierte Prozente
Gültig 18 13 ,2 ,2 ,2
19 42 ,7 ,7 ,9
20 54 ,8 ,8 1,7
21 57 ,9 ,9 2,6
22 65 1,0 1,0 3,6
23 69 1,1 1,1 4,7
24 111 1,7 1,7 6,4
25 121 1,9 1,9 8,3
26 138 2,2 2,2 10,5
27 114 1,8 1,8 12,3
28 162 2,5 2,5 14,8
29 156 2,4 2,4 17,2

Hauptfenster und aktives Fenster

Wenn Sie mehr als ein Viewer-Fenster geöffnet haben, werden Ausgaben an das **Hauptfenster** des Viewers geleitet. Wenn Sie mehr als ein Fenster für den Syntax-Editor geöffnet haben, wird Syntax in das Hauptfenster des Syntax-Editors eingefügt. Die Hauptfenster weisen ein Pluszeichen in der Titelleiste auf. Sie können die Hauptfenster jederzeit wechseln.

Das Hauptfenster darf nicht mit dem **aktiven** Fenster verwechselt werden. Das aktive Fenster ist das aktuell ausgewählte Fenster. Wenn sich mehrere Fenster überlappen, wird das aktive Fenster im Vordergrund angezeigt. Wenn Sie ein neues Fenster öffnen, wird dieses automatisch sowohl das aktive Fenster als auch das Hauptfenster.

Wechseln des Hauptfensters

- Aktivieren Sie das Fenster, welches das Hauptfenster werden soll. Klicken Sie dazu auf eine beliebige Stelle im Fenster.
- Klicken Sie auf die Schaltfläche “Hauptfenster” in der Symbolleiste (das Symbol mit dem Pluszeichen).

  oder

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Extras
  Hauptfenster

Anmerkung: Im Daten-Editor wird durch das aktive Daten-Editor-Fenster festgelegt, welches Daten-Set in Berechnungen oder Analysen verwendet wird. Es gibt kein “Hauptfenster” im Daten-Editor. Für weitere Informationen siehe “Grundsätzlicher Umgang mit mehreren Datenquellen” in Kapitel 6 auf S. 122.

Menüs

Viele der Aufgaben, die Sie in SPSS ausführen können, stehen nach Auswahl der entsprechenden Menübefehle zur Verfügung. Jedes Fenster in SPSS verfügt über eine eigene Menüleiste mit entsprechenden Menüs für diese Art von Fenster.

Die Menüs “Analysieren” und “Grafiken” sind in allen Fenstern verfügbar. Dies vereinfacht das Erstellen von Ausgaben, ohne daß Sie dabei zwischen Fenstern wechseln müssen.

Statusleiste

Die Statusleiste am unteren Rand jedes Fensters von SPSS enthält die folgenden Informationen:

Befehlsstatus. Bei jeder ausgeführten Prozedur und jedem ausgeführten Befehl wird angezeigt, wie viele Fälle bereits verarbeitet wurden. Bei statistischen Prozeduren mit iterativer Verarbeitung wird die Anzahl der Iterationen angezeigt.

Filterstatus. Wenn Sie eine zufällige Stichprobe oder eine Teilmenge der Fälle zum Analysieren ausgewählt haben, zeigt Ihnen die Meldung Filter an, daß die Fälle gefiltert und somit nicht alle Fälle in der Datendatei bei der Analyse berücksichtigt werden.

Gewichtungsstatus. Die Meldung Gewichtung an zeigt an, daß die Fälle mit einer Gewichtungsvariablen analysiert werden.

Aufspaltungsstatus der Datei. Die Meldung Datei aufteilen an bedeutet, daß die Datendatei zur Analyse anhand der Werte von einer oder mehr Gruppenvariablen in verschiedene Gruppen aufgeteilt wurde.

Ein- und Ausblenden der Statusleiste

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Ansicht
  Statusleiste

Dialogfelder

Wenn Sie eine Option aus einem Menü auswählen, wird in den meisten Fällen ein Dialogfeld geöffnet. Sie verwenden die Dialogfelder zum Auswählen von Variablen und Optionen für die Analyse.

Dialogfelder für statistische Prozeduren und Diagramme weisen in der Regel folgende zwei Hauptkomponenten auf:

Liste der Quellvariablen. Eine Variablenliste in der Arbeitsdatei. In der Quelliste werden nur die Typen von Variablen angezeigt, die für die ausgewählte Prozedur zulässig sind. Kurze und lange String-Variablen können in vielen Prozeduren nur eingeschränkt verwendet werden.

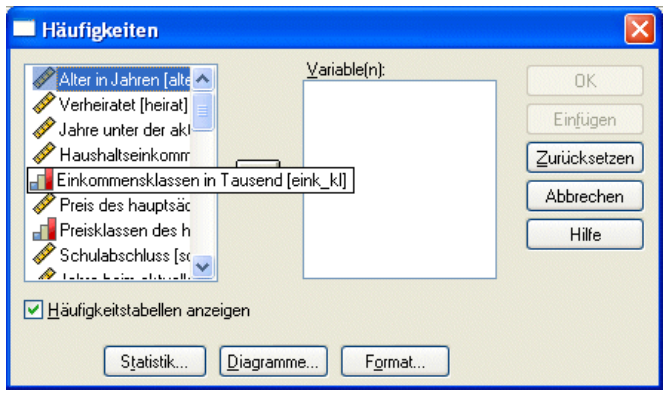
Zielvariablenliste(n). Hierbei handelt es sich um eine oder mehrere Listen mit den Variablen, die Sie zur Analyse ausgewählt haben, also beispielsweise die Liste mit den abhängigen und die Liste mit den unabhängigen Variablen.

Variablennamen und Variablenlabels in Listen von Dialogfeldern

In Listen von Dialogfeldern können Sie entweder Variablennamen oder Variablenlabels anzeigen lassen.

- Ob Variablennamen oder -beschriftungen angezeigt werden sollen, können Sie in allen Fenstern mit dem Befehl Optionen im Menü “Bearbeiten” festlegen.
- Variablenlabels können in der Variablenansicht des Daten-Editors erstellt oder bearbeitet werden.
- Für Daten, die aus Datenbankquellen importiert wurden, werden Feldnamen als Variablenlabels verwendet.
- Bei langen Labels können Sie den Mauszeiger über das Label in der Liste halten, um das gesamte Label anzeigen zu lassen.
- Falls kein Variablenlabel definiert ist, wird der Variablenname angezeigt.

Abbildung 1-2
In Dialogfeldern angezeigte Variablenlabels



Steuerelemente in Dialogfeldern

Die meisten Dialogfelder enthalten fünf Steuerelemente:

OK. Hiermit führen Sie die Prozedur aus. Klicken Sie nach dem Auswählen von Variablen und dem Festlegen von zusätzlichen Angaben auf OK, um die Prozedur auszuführen und das Dialogfeld zu schließen.

Einfügen. Hiermit erzeugen Sie aus den Einstellungen im Dialogfeld Befehlssyntax und fügen diese Syntax in ein Syntax-Fenster ein. Sie können die Befehle um zusätzliche Funktionen erweitern, auf die Sie sonst nicht über die Dialogfelder zugreifen können.

Zurücksetzen. Hiermit heben Sie die Auswahl von Variablen in den Listen der ausgewählten Variablen auf und setzen alle Einstellungen im aktuellen Dialogfeld und in allen untergeordneten Dialogfeldern auf die Standardeinstellungen zurück.

Abbrechen. Hiermit verwerfen Sie alle Änderungen, die seit dem letzten Öffnen an den Einstellungen im Dialogfeld vorgenommen wurden, und schließen das Dialogfeld. Innerhalb einer Sitzung bleiben die Einstellungen in einem Dialogfeld bestehen. Die Einstellungen in einem Dialogfeld werden beibehalten, bis Sie diese überschreiben.

Hilfe. Stellt kontextsensitive Hilfe bereit. Hiermit wechseln Sie zu einem Hilfefenster mit Informationen zum aktuellen Dialogfeld. Hilfe zu einzelnen Steuerelementen in Dialogfeldern erhalten Sie außerdem, indem Sie mit der rechten Maustaste auf das entsprechende Steuerelement klicken.

Untergeordnete Dialogfelder

Da die meisten Prozeduren sehr flexibel anwendbar sind, konnten nicht alle Auswahlmöglichkeiten in jeweils einem einzigen Dialogfeld untergebracht werden. In der Regel enthält das Hauptdialogfeld die Informationen, die mindestens zum Ausführen einer Prozedur erforderlich sind. Zusätzliche Angaben erfolgen in untergeordneten Dialogfeldern.

Drei Punkte (...) nach dem Namen eines Steuerelements im Hauptdialogfeld zeigen an, daß hiermit ein untergeordnetes Dialogfeld geöffnet werden kann.

Auswählen von Variablen

Zum Auswählen einer einzelnen Variablen markieren Sie diese in der Liste der Quellvariablen und klicken dann auf die Pfeilschaltfläche neben dem Listenfeld mit den Zielvariablen. Wenn nur eine Liste für Zielvariablen vorhanden ist, können Sie auf einzelne Variablen doppelklicken, um diese aus der Liste der Quellvariablen in die Liste der Zielvariablen zu verschieben.

Sie können auch mehrere Variablen gleichzeitig auswählen:

- Um mehrere Variablen auszuwählen, die nacheinander in der Variablenliste stehen, klicken Sie auf die erste Variable, halten die Umschalttaste gedrückt und klicken anschließend auf die letzte Variable in der Gruppe.
- Um mehrere Variablen auszuwählen, die nicht nacheinander in der Variablenliste stehen, klicken Sie auf die erste Variable und halten beim Klicken auf die weiteren Variablen die Strg-Taste gedrückt.

Symbole für Variablenlisten

Die Symbole, die neben den Variablen in Dialogfeldern angezeigt werden, liefern Informationen über den Variablentyp und das Messniveau.

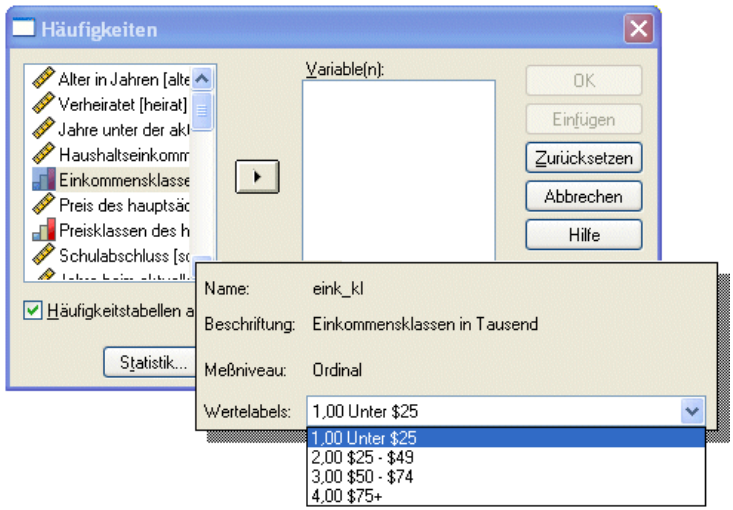
Messniveau	**Datentyp**			
	Numerisch	String	Datum	Zeit
Metrisch		entfällt		
Ordinal				
Nominal				

Aufrufen von Informationen zu Variablen in einem Dialogfeld

▶ Klicken Sie mit der rechten Maustaste auf eine Variable in der Liste der Quell- oder Zielvariablen.

▶ Wählen Sie Info zu Variable aus.

Abbildung 1-3
Variablenbeschreibung



Grundlegende Schritte bei der Datenanalyse

Das Analysieren von Daten in SPSS ist sehr einfach. Sie müssen nur die folgenden Schritte ausführen:

Übertragen Ihrer Daten in SPSS. Sie können eine bereits gespeicherte SPSS-Datendatei öffnen; eine Tabellenblatt-, Datenbank- oder Text-Datendatei einlesen oder die Daten direkt im Daten-Editor eingeben.

Auswählen einer Prozedur. Wählen Sie eine Prozedur aus den Menüs aus, um Statistiken zu berechnen oder ein Diagramm zu erstellen.

Auswählen der Variablen für die Analyse. Die Variablen in der Datendatei werden in einem Dialogfeld für die Prozedur angezeigt.

Führen Sie die Prozedur aus, und betrachten Sie die Ergebnisse. Die Ergebnisse werden im Viewer angezeigt.

Statistik-Assistent

Wenn Ihnen SPSS oder die in SPSS verfügbaren statistischen Prozeduren nicht vertraut sind, kann Ihnen der Statistik-Assistent den Einstieg erleichtern. Einfache Fragen in einer verständlichen Sprache und visuelle Beispiele helfen Ihnen bei der Auswahl der grundlegenden statistischen Funktionen und Diagrammfunktionen, die sich am besten für Ihre Daten eignen.

Zum Aufrufen des Statistik-Assistenten wählen Sie die folgenden Befehle aus den Menüs eines beliebigen Fensters in SPSS aus:

Hilfe
  Statistik-Assistent

Der Statistik-Assistent enthält nur eine ausgewählte Teilmenge der Prozeduren aus dem Base-System von SPSS. Er dient der allgemeinen Unterstützung bei vielen grundlegenden und häufig verwendeten statistischen Verfahren.

Weitere Informationen zu SPSS

Eine umfassende Übersicht über die Grundlagen von SPSS finden Sie im Online-Lernprogramm. Wählen Sie folgende Befehle aus den Menüs eines beliebigen Fensters von SPSS aus:

Hilfe
  Lernprogramm

Kapitel
2

Aufrufen der Hilfe

Hilfestellung ist auf verschiedene Weise verfügbar:

Menü "Hilfe". Das Menü "Hilfe" in den meisten SPSS-Fenstern führt zum Haupthilfesystem, außerdem zu Lernprogrammen und technischen Referenzen.

- **Themen**. Hiermit können Sie auf die Registerkarten "Inhalt", "Index" und "Suchen" zugreifen. Verwenden Sie diese Registerkarten bei der Suche nach bestimmten Hilfethemen.
- **Lernprogramm.** Illustrierte, schrittweise Anleitungen für die Verwendung zahlreicher Grundfunktionen in SPSS. Es ist nicht notwendig, das gesamte Lernprogramm vom Anfang bis zum Ende durchzuarbeiten. Sie können die gewünschten Themen direkt auswählen, nach Wunsch zwischen den Themen wechseln, die Themen in beliebiger Reihenfolge abrufen oder auch bestimmte Themen über den Index oder das Inhaltsverzeichnis suchen.
- **Fallstudien.** Praktische Beispiele für die Erstellung verschiedener Arten von statistischen Analysen und für die Interpretation der Ergebnisse. Die in den Beispielen verwendeten Datendateien werden auch bereitgestellt. Sie können also die Beispiele durcharbeiten, um zu verfolgen, wie die Ergebnisse zustande kommen. Sie können die gewünschten Prozeduren im Inhaltsverzeichnis auswählen oder nach relevanten Themen im Index suchen.
- **Statistik-Assistent.** Dieser Assistent unterstützt Sie bei der Suche nach der Prozedur, die Sie verwenden möchten. Nachdem Sie Ihre Auswahl getroffen haben, öffnet der Statistik-Assistent das Dialogfeld für die Statistik-, Berichts- oder Diagrammprozedur, das die ausgewählten Kriterien erfüllt. Der Statistik-Assistent bietet Zugang zu den meisten Statistik- und Berichtsprozeduren sowie auf einen Großteil der Diagrammprozeduren im Base-System.

- **Befehlssyntax-Referenz (Command Syntax Reference).** Detaillierte Informationen zur Befehlssyntax sind auf zwei Arten verfügbar: als Bestandteil der umfassenden Hilfesystems und als separates Dokument im PDF-Format im Handbuch *SPSS Command Syntax Reference*, das auch über das Menü "Hilfe" verfügbar ist.
- **Statistische Algorithmen.** Die meisten der für die statistischen Prozeduren verwendeten Algorithmen sind im PDF-Format über das Menü "Hilfe" sowie aus den Hilfethemen für die enstprechenden Dialogfelder verfügbar.

Kontextsensitive Hilfe. An zahlreichen Stellen der Benutzeroberfläche können Sie kontextsensitive Hilfe abrufen.

- **Schaltflächen für Hilfe in Dialogfeldern.**Die meisten Dialogfelder verfügen über die Schaltfläche "Hilfe", mit der Sie das entsprechende Hilfethema für das Dialogfeld direkt aufrufen können. In diesem Hilfethema finden Sie allgemeine Informationen und Verknüpfungen zu verwandten Themen.
- **Hilfe zu Dialogfeldern über das Kontextmenü.** Viele Dialogfelder bieten kontextsensitive Hilfe für einzelne Steuerungen und Funktionen. Wenn Sie mit der rechten Maustaste auf ein Steuerelement in einem Dialogfeld klicken und dann Direkthilfe aus dem Kontextmenü auswählen, erhalten Sie eine Beschreibung des Steuerelements und seiner Verwendung. (Falls im Kontextmenü nicht der Eintrag Direkthilfe aufgeführt wird, steht diese Art der Hilfe im betreffenden Dialogfeld nicht zur Verfügung.)
- **Hilfe zu Pivot-Tabellen über das Kontextmenü.** Wenn Sie mit der rechten Maustaste auf Begriffe einer im Viewer aktivierten Pivot-Tabelle klicken und dann Direkthilfe aus dem Kontextmenü auswählen, erhalten Sie eine Definition dieser Begriffe.
- **Fallstudien.** Klicken Sie mit der rechten Maustaste auf eine Pivot-Tabelle, und wählen Sie im Kontextmenü den Befehl Fallstudien. Sie gelangen direkt zu einem ausführlichen Beispiel für die Prozedur, mit der die betreffende Tabelle erzeugt wurde. (Falls im Kontextmenü nicht der Eintrag Fallstudien aufgeführt wird, steht diese Art der Hilfe für die vorliegende Prozedur nicht zur Verfügung.)
- **Befehlssyntax.** Zeigen Sie in einem Befehlssyntaxfenster auf eine beliebige Position innerhalb eines Syntaxblocks für einen Befehl, und drücken Sie F1 auf der Tastatur. Das vollständige Befehlssyntaxdiagramm für diesen Befehl wird eingeblendet. Die vollständige Dokumentation für die Befehlssynatx ist über die

Verknüpfungen in den Listen der verwandten Themen und auf der Registerkarte "Inhalt" der Hilfe verfügbar.

- **Skripts und OLE-Automatisierung.** Im Skript-Fenster (Menü "Datei", Befehl "Neu" oder "Öffnen", "Skript") erhalten Sie über das Menü "Hilfe" Informationen zur Skriptsprache und zu den Objekten, Methoden und Eigenschaften für die OLE-Automatisierung in SPSS. Kontextsensitive Hilfe können Sie im Skriptfenster mit F1 oder F2 (Objektkatalog) aufrufen.

Einstellungen für Microsoft Internet Explorer

Bei den meisten Funktionen der Hilfe in dieser Anwendung werden Funktionen von Microsoft Internet Explorer verwendet. In der Standardeinstellung einiger Versionen von Internet Explorer (einschließlich der Version, die in Microsoft XP Service Pack 2 enthalten ist) werden so genannte aktive Inhalte im Fenster von Internet Explorer auf Ihrem lokalen Computer unterdrückt. Diese Standardeinstellung kann dazu führen, dass einige Inhalte der Hilfe unterdrückt werden. Um Zugriff auf die gesamte Hilfe zu erlangen, können Sie die Standardeinstellung in Internet Explorer ändern.

▶ Wählen Sie in Internet Explorer die folgenden Menübefehle aus:
Extras
Internetoptionen...

▶ Klicken Sie auf die Registerkarte Erweitert.

▶ Führen Sie in der Liste einen Bildlauf bis zum Abschnitt Sicherheit durch.

▶ Aktivieren Sie Ausführung aktiver Inhalte in Dateien auf dem lokalen Computer zulassen.

Sonstige Ressourcen

Diese Ressourcen können Ihnen weiterhelfen, wenn Sie die gesuchten Informationen nicht im Hilfesystem finden:

- **Entwicklerhandbuch für SPSS für Windows.** Hier finden Sie Informationen und Beispiele für die Entwickler-Tools in SPSS für Windows, z. B. zur OLE-Automatisierung, zu APIs von Drittanbietern, zur Ein-/Ausgabe-DLL, zum

Produktionsmodus und zur Skriptfunktion. Das *Entwicklerhandbuch* befindet sich im PDF-Format im Verzeichnis *SPSS\developer* der Installations-CD.

- **Website des technischen Supports.** Antworten auf viele häufig auftretende Probleme finden Sie unter *http://support.spss.com*. (Für die Website des technischen Supports benötigen Sie eine Anmelde-ID und ein Passwort. Weitere Informationen zum Anfordern einer ID und eines Paßworts finden Sie unter der genannten URL.)

Verwenden des Inhaltsverzeichnisses der SPSS-Hilfe

▶ Wählen Sie in einem beliebigen Fenster die folgenden Befehle aus den Menüs aus:

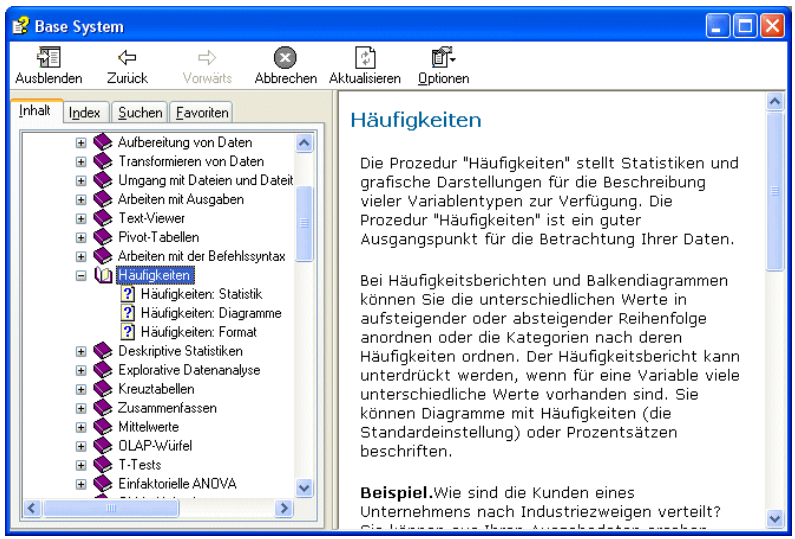
Hilfe
  Themen

▶ Klicken Sie auf die Registerkarte Inhalt.

▶ Wenn Sie auf Einträge mit einem Buchsymbol doppelklicken, können Sie den Inhalt ein- oder ausblenden.

▶ Wenn Sie auf einen Eintrag klicken, wird das entsprechende Hilfethema aufgerufen.

Abbildung 2-1
Hilfe-Fenster mit angezeigter Registerkarte "Inhalt"



Verwenden des Hilfe-Indexes

- Wählen Sie in einem beliebigen Fenster die folgenden Befehle aus den Menüs aus:

  Hilfe
  Themen

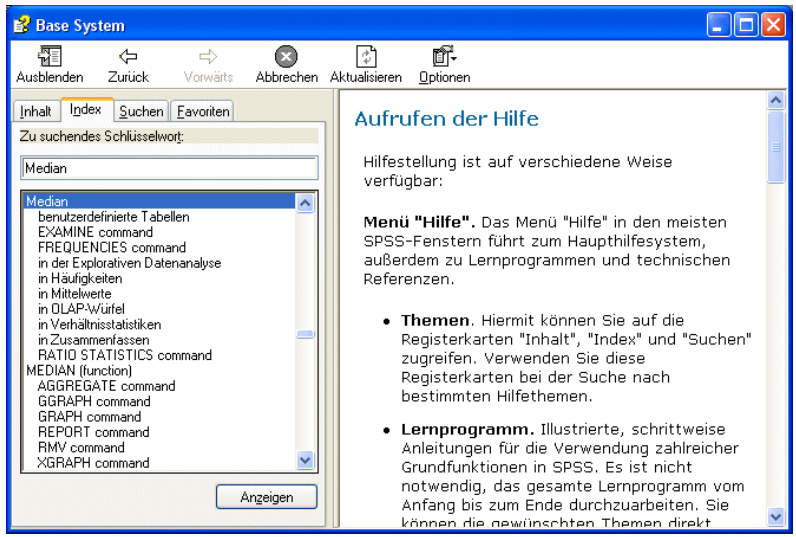
- Klicken Sie auf die Registerkarte Index.

- Geben Sie im Index den Suchbegriff ein.

- Doppelklicken Sie auf das gewünschte Thema.

  Im Hilfe-Index erfolgt die Suche nach dem eingegebenen Text schrittweise, wobei die bestmögliche Übereinstimmung mit dem Index ausgewählt wird.

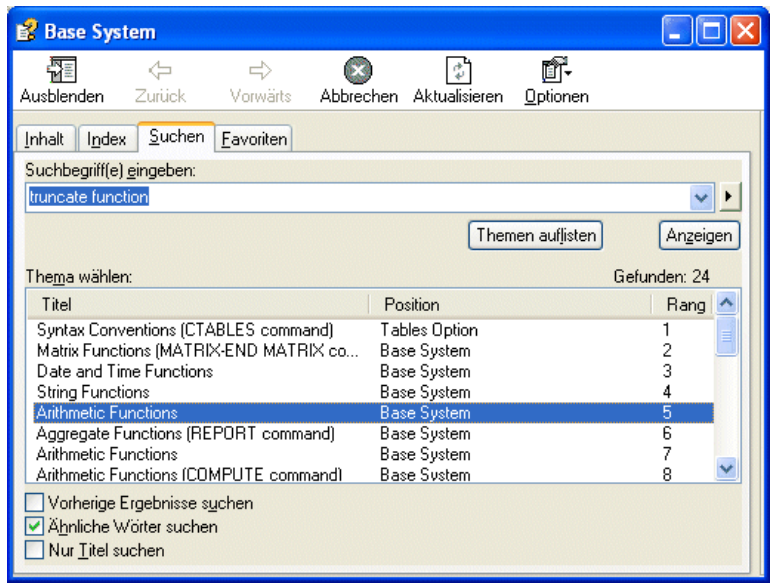
Abbildung 2-2
Registerkarte "Index" und schrittweise Suche



Verwenden der Suche in der Hilfe

Auf der Registerkarte "Suche" können Sie eine Volltextsuche in den Thementiteln, den Hilfethemen selbsr und in den Indextexten durchführen. Die Themen werden nach der Häufigkeit angeordnet, mit welcher der Suchbegriff in den Themen bzw. dem entsprechenden Index auftritt.

Abbildung 2-3
Registerkarte "Suchen" der Hilfe



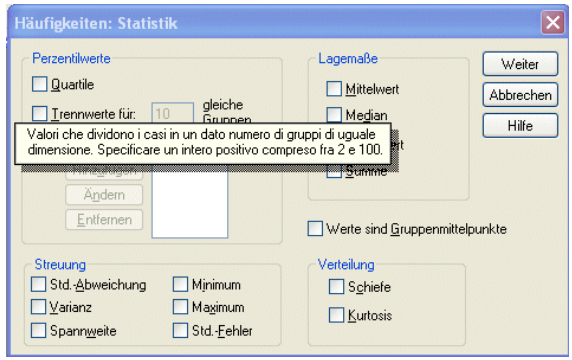
- Die Vollextsuche ist am effektivsten, wenn Sie die Suchkriterien verfeinern, um die Anzahl der Themen in der Liste zu begrenzen. Wenn Sie zum Beispiel wissen, dass Sie eine Funktion zum Stutzen von numerischen Werten suchen, ergibt der Suchbegriff "Stutzen Funktion" eine bessere Liste von möglichen Themen als der alleinige Suchbegriff "Stutzen".
- Wenn bei einem Hilfethema in Klammern ein Befehlsname aufgeführt wird, bezieht sich dieses Thema auf die Befehlssyntax. Themen ohne einen Befehlsnamen in Klammern stellen im Allgemeinen Hilfethemen für Dialogfelder und andere Themen der grafischen Benutzeroberfläche dar.

Aufrufen der Hilfe zu Steuerelementen in Dialogfeldern

- ▶ Klicken Sie in einem Dialogfeld mit der rechten Maustaste auf das Steuerelement, zu dem Sie Informationen benötigen.
- ▶ Wählen Sie aus dem Kontextmenü Direkthilfe aus.

In einem Popup-Fenster finden Sie eine Beschreibung der Schaltfläche und Hinweise zu deren Verwendung. Allgemeine Informationen zu einem Dialogfeld können Sie über die zugehörige Schaltfläche “Hilfe” abrufen.

Abbildung 2-4
Hilfe zu Schaltflächen in Dialogfeldern über die rechte Maustaste

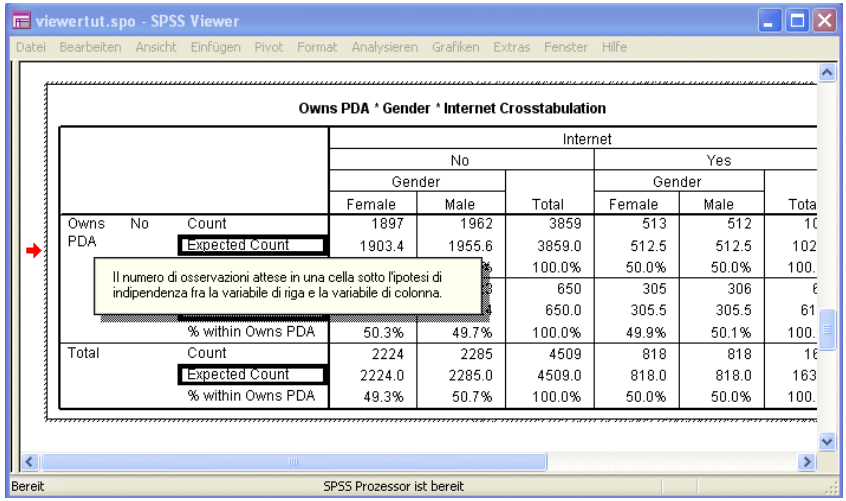


Aufrufen der Hilfe zu ausgegebenen Begriffen

- Doppelklicken Sie auf die Pivot-Tabelle, um diese zu aktivieren.
- Klicken Sie mit der rechten Maustaste auf den Term, zu dem Sie Erklärungen benötigen.
- Wählen Sie Direkthilfe aus dem Kontextmenü aus.

  In einem Popup-Fenster wird eine Definition des Terms angezeigt.

Abbildung 2-5
Glossar/Hilfe zu aktiver Pivot-Tabelle über die rechte Maustaste



Verwenden von Fallstudien

- Klicken Sie im Viewer-Fenster mit der rechten Maustaste auf eine Pivot-Tabelle.
- Wählen Sie im Kontextmenü die Option Fallstudien aus.

Kopieren von Hilfetext aus einem Popup-Fenster

- Klicken Sie mit der rechten Maustaste in das Popup-Fenster.
- Wählen Sie im Kontextmenü Kopieren.

  Der vollständige Text aus dem Popup-Fenster wird kopiert.

Kapitel
3

Datendateien

Datendateien können in einer Reihe von verschiedenen Formaten vorliegen. SPSS kann mit vielen dieser Formaten arbeiten, unter anderem mit den folgenden:

- Tabellenkalkulationsblätter aus Excel und Lotus
- Datenbanktabellen aus vielen Datenbankquellen, einschließlich Oracle, SQL-Server, Access, DBASE und andere
- Mit Tabulatoren als Trennzeichen versehene und andere Typen von Textdateien
- Datendateien im SPSS-Format, die unter anderen Betriebssystemen erstellt wurden,
- SYSTAT-Datendateien
- SAS-Datendateien
- Stata-Datendateien

Öffnen einer Datendatei

Neben den im SPSS-Format gespeicherten Dateien lassen sich die Dateien von Excel, SAS und Stata sowie mit Tabulatoren als Trennzeichen versehene und andere Dateien öffnen, ohne diese Dateien in ein Zwischenformat umzuwandeln oder Datendefinitionseingaben vorzunehmen.

So öffnen Sie Datendateien:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Datei
  Öffnen
    Daten...

▶ Wählen Sie im Dialogfeld “Datei öffnen” die zu öffnende Datei aus.

▶ Klicken Sie auf Öffnen.

Die folgenden Optionen sind verfügbar:

- Aus der ersten Zeile von Tabellenkalkulationsdateien oder Dateien, in denen Tabulatoren als Trennzeichen eingesetzt werden, können die Namen von Variablen eingelesen werden.
- Bei Tabellenkalkulationsdateien kann der Zellenbereich angegeben werden, der eingelesen werden soll.
- Geben Sie an, welches Arbeitsblatt aus der Excel-Datei gelesen werden soll (Excel 5 oder Nachfolgeversionen).

Datendateitypen

SPSS. Hiermit werden Datendateien geöffnet, die mit SPSS für Windows, Macintosh oder UNIX sowie mit dem DOS-Produkt SPSS/PC+ gespeichert wurden.

SPSS/PC+. Hiermit werden Datendateien von SPSS/PC+ geöffnet.

SYSTAT. Hiermit werden Datendateien von SYSTAT geöffnet.

SPSS Portable. Hiermit werden SPSS-Datendateien geöffnet, die im portablen Format gespeichert wurden. Das Speichern einer Datei im portablen Format erfordert erheblich mehr Zeit als das Speichern der Datei im SPSS-Format.

Excel. Hiermit werden Excel-Dateien geöffnet.

Lotus 1-2-3. Hiermit werden Datendateien geöffnet, die im Format von Lotus 1-2-3, den Versionen 3.0, 2.0 oder der Version 1A gespeichert wurden.

SYLK. Hiermit werden die im SYLK-Format gespeicherten Datendateien geöffnet. Dieses Format wird bei manchen Tabellenkalkulationsprogrammen eingesetzt.

dBASE. Hiermit werden dBASE-Dateien im Format von dBASE IV, dBASE III oder III PLUS sowie dBASE II geöffnet. Jeder Fall ist eine Zeile im Datensatz. Beim Speichern einer Datei in diesem Format gehen die Bezeichnungen von Variablen, die Wertelabels und die Angaben zu fehlenden Werten verloren.

SAS Long File Name. SAS Versionen 7-9 für Windows, lange Erweiterung.

SAS Short File Name. SAS Versionen 7-9 für Windows, kurze Erweiterung.

SAS v6 für Windows. SAS Version 6.08 für Windows und OS2.

SAS v6 für UNIX. SAS Version 6 für UNIX (Sun, HP, IBM).

SAS Transport. SAS-Transportdatei.

Stata. Stata Versionen 4-8.

Text. ASCII-Textdatei.

Datei öffnen: Optionen

Variablennamen lesen. Bei Tabellenkalkulationsdateien können die Namen der Variablen aus der ersten Zeile der Datei oder des angegebenen Bereichs eingelesen werden. Die Werte werden nach Bedarf umgewandelt, um gültige Variablennamen zu erstellen. Dabei werden Leerzeichen in Unterstriche umgewandelt.

Arbeitsblatt. Dateien von Excel 5 oder Nachfolgeversionen können mehrere Arbeitsblätter enthalten. In der Standardeinstellung liest der Daten-Editor das erste Arbeitsblatt. Wenn Sie ein anderes Arbeitsblatt einlesen möchten, wählen Sie es aus der Dropdown-Liste aus.

Bereich. Bei Datendateien aus Tabellenkalkulationen ist es außerdem möglich, nur einen bestimmten Zellenbereich einzulesen. Verwenden Sie beim Festlegen des Zellenbereichs dieselbe Methode wie im Tabellenkalkulationsprogramm.

Einlesen von Dateien aus Excel 5 oder nachfolgenden Versionen

Beim Einlesen von Dateien aus Excel 5 oder nachfolgenden Versionen gelten die folgenden Regeln:

Datentypen und Breiten. Jede Spalte stellt eine Variable dar. Der Datentyp und die Breite jeder Variablen werden durch den Datentyp und die Breite in der Excel-Datei bestimmt. Wenn eine Spalte mehr als einen Datentyp enthält (beispielsweise Datumsangaben und numerische Daten), wird der Datentyp auf “String” gesetzt, und alle Werte werden als gültige String-Werte eingelesen.

Leere Zellen. Bei numerischen Variablen werden die leeren Zellen in systemdefinierte fehlende Werte konvertiert. Diese werden durch einen Punkt dargestellt. Bei String-Variablen stellen leere Zellen gültige Werte dar. Leere Zellen werden daher als gültige String-Variablen behandelt.

Variablennamen. Wenn aus der ersten Zeile der Excel-Datei (oder der ersten Zeile des angegebenen Bereichs) Variablennamen eingelesen werden, werden Werte, die nicht den Regeln für Variablennamen entsprechen, in gültige Variablennamen umgewandelt, und die ursprünglichen Spaltenüberschriften werden als Variablenlabels gespeichert. Falls keine Variablennamen aus der Excel-Datei eingelesen werden, erhalten die Variablen Standardnamen.

Einlesen von älteren Excel-Dateien und anderen Tabellenkalkulationsdateien

Beim Einlesen von Dateien von älteren Versionen von Excel und anderen Tabellenkalkulationsprogrammen gelten die folgenden Regeln:

Datentypen und Breiten. Die Datentypen und Breiten der Variablen werden durch die Datentypen und Spaltenbreiten der ersten Zellen mit Daten in den Spalten festgelegt. Werte anderer Typen werden in systemdefiniert fehlende Werte konvertiert. Wenn die erste Zelle mit Daten in einer Spalte leer ist, wird der globale Standarddatentyp des Tabellenkalkulationsblatts verwendet. In der Regel handelt es sich hierbei um einen numerischen Datentyp.

Leere Zellen. Bei numerischen Variablen werden die leeren Zellen in systemdefinierte fehlende Werte konvertiert. Diese werden durch einen Punkt dargestellt. Bei String-Variablen stellen leere Zellen gültige Werte dar. Leere Zellen werden daher als gültige String-Variablen behandelt.

Variablennamen. Wenn Sie die Namen der Variablen nicht aus dem Tabellenkalkulationsblatt einlesen, verwendet SPSS bei Excel- und Lotus-Dateien Buchstaben für die Bezeichnung der Spalten, also *A*, *B*, *C* usw., als Variablennamen. Bei SYLK-Dateien und Excel-Dateien, die im Anzeigeformat "R1C1" gespeichert wurden, verwendet SPSS den Buchstaben *C* und die Spaltennummer, also *C1*, *C2*, *C3* usw., als Variablennamen.

Einlesen von dBASE-Dateien

Datenbankdateien sind bezüglich der Logik ähnlich wie Datendateien im SPSS-Format aufgebaut. Bei dBASE-Dateien gelten die folgenden allgemeinen Regeln:

- Feldnamen werden in gültige Variablennamen umgewandelt.

- Falls in den Feldnamen in dBASE Doppelpunkte verwendet werden, werden diese in Unterstriche umgewandelt.
- Datensätze, die zwar zum Löschen markiert, aber noch nicht bereinigt wurden, werden berücksichtigt. SPSS erstellt eine neue String-Variable, *D_R*, und weist dieser bei Fällen, die zum Löschen markiert wurden, ein Sternchen zu.

Einlesen von Stata-Dateien

Bei Stata-Datendateien gelten die folgenden allgemeinen Regeln:

- **Variablennamen.** Stata-Variablennamen werden unter der Berücksichtigung der Groß- und Kleinschreibung in SPSS-Variablennamen umgewandelt. Stata-Variablennamen, die bis auf die Groß- und Kleinschreibung übereinstimmen, werden in gültige SPSS-Variablennamen umgewandelt, indem ein Unterstrich und fortlaufende Buchstaben angehängt werden (*_A*, *_B*, *_C*, ..., *_Z*, *_AA*, *_AB*, ... usw.).
- **Variablenlabels.** Stata-Variablenlabels werden in SPSS-Variablenlabels umgewandelt.
- **Wertelabels.** Mit Ausnahme der Stata-Wertelabels, die "erweiterten" fehlenden Werten zugewiesen sind, werden Stata-Wertelabels in SPSS-Wertelabels umgewandelt.
- **Fehlende Werte.** "Erweiterte" fehlende Werte aus Stata werden in systemdefinierte fehlende Werte umgewandelt.
- **Umwandlung von Daten.** Stata-Werte im Datumsformat werden in Werte mit dem SPSS-Format `DATE` (T_m_J) umgewandelt. Stata-Datumswerte für Zeitreihen (Wochen, Monate, Quartale usw.) werden in das einfache numerische Format (F) umgewandelt. Dabei bleibt der ursprüngliche, ganzzahlige Wert erhalten. Dies ist die Anzahl an Wochen, Monaten, Quartale usw. seit dem Beginn des Jahres 1960.

Einlesen von Datenbankdateien

SPSS kann Daten aus Datenbankdateien in beliebigen Formaten einlesen, wenn Sie über einen entsprechenden Datenbanktreiber verfügen. Im Modus für lokale Analysen müssen die erforderlichen Treiber auf dem lokalen Computer installiert sein. Im Modus für verteilte Analysen (verfügbar mit SPSS Server) müssen die Treiber auf

dem Remote-Server installiert sein. Für weitere Informationen siehe “Modus für verteilte Analysen” in Kapitel 4 auf S. 83.

So lesen Sie Datenbankdateien ein

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Datei
  Datenbank öffnen
  Neue Abfrage...

- Wählen Sie die Datenquelle aus.
- Falls erforderlich (abhängig von der Datenquelle), wählen Sie die Datenbankdatei aus, und/oder geben Sie einen Anmeldenamen, ein Passwort und andere Informationen ein.
- Wählen Sie die Tabelle(n) und Felder aus. (Bei OLE DB-Datenquellen können Sie nur eine Tabelle auswählen.)
- Legen Sie gegebenenfalls Relationen zwischen den Tabellen fest.
- Die folgenden Optionen sind verfügbar:
  - Auswahlkriterien für die Daten festlegen,
  - Eine Aufforderung für benutzerdefinierte Eingaben hinzufügen, um eine Parameterabfrage zu erstellen,
  - Speichern Sie die erstellte Abfrage, bevor Sie sie ausführen.

So können Sie gespeicherte Datenbankabfragen bearbeiten

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Datei
  Datenbank öffnen
  Abfrage bearbeiten...

- Wählen Sie die Abfragedatei (**.spq*) aus, die Sie bearbeiten möchten.
- Zum Erstellen einer neuen Abfrage folgen Sie den Anweisungen.

So lesen Sie Datenbankdateien mit gespeicherten Abfragen ein

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Datei
  Datenbank öffnen
  Abfrage ausführen...

- Wählen Sie die auszuführende Abfragedatei (**.spq*) aus.

- Falls erforderlich (abhängig von der Datenbankdatei), geben Sie einen Anmeldenamen und ein Passwort ein.

- Wenn für die Abfrage eine Eingabeaufforderung definiert wurde, müssen Sie ggf. weitere Informationen eingeben (beispielsweise das Quartal, für das Sie die Verkaufszahlen abrufen möchten).

Auswählen einer Datenquelle

Wählen Sie im ersten Bildschirm des Datenbank-Assistenten den Typ der einzulesenden Datenquelle aus.

ODBC-Datenquellen

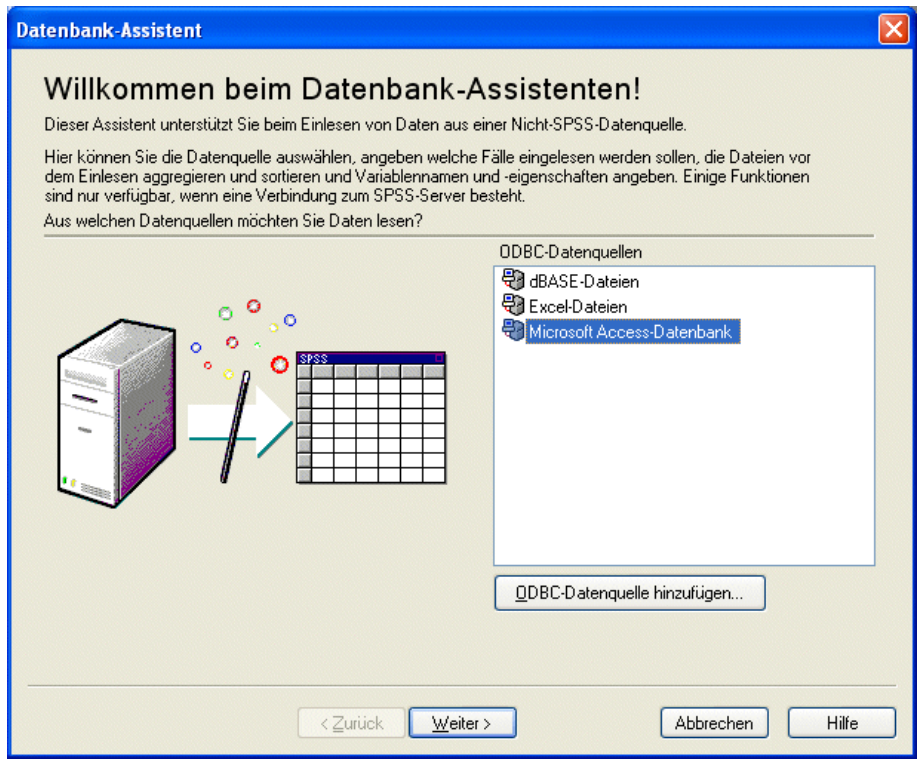
Wenn Sie noch keine ODBC-Datenquelle konfiguriert haben oder eine neue Datenquelle hinzufügen möchten, klicken Sie auf ODBC-Datenquelle hinzufügen. Im Modus für verteilte Analysen (verfügbar mit SPSS Server) steht diese Schaltfläche nicht zur Verfügung. Wenn Sie Datenquellen im Modus für verteilte Analysen hinzufügen möchten, wenden Sie sich an Ihren Systemadministrator.

Eine ODBC-Datenquelle besteht aus zwei wichtigen Informationen: dem Treiber, der zum Zugreifen auf die Daten verwendet wird, und dem Speicherort der Datenbank, auf die Sie zugreifen möchten. Wenn Sie Datenquellen definieren

möchten, muß der entsprechende Treiber installiert sein. Für den Modus für lokale Analysen können Sie Treiber von der CD-ROM dieses Programms installieren:

- **SPSS Data Access Pack.** Hiermit werden Treiber für verschiedenen Datenbankformate installiert. Diese Funktion ist im automatisch angezeigten Menü der CD verfügbar.
- **Microsoft Data Access Pack.** Hiermit werden Treiber für Programme von Microsoft (beispielsweise Microsoft Access) installiert. Um das Microsoft Data Access Pack zu installieren, doppelklicken Sie im Ordner "Microsoft Data Access Pack" auf der CD-ROM auf Microsoft Data Access Pack.

Abbildung 3-1
Dialogfeld "Datenbank-Assistent"


Datenbank-Assistent
Willkommen beim Datenbank-Assistenten!
Dieser Assistent unterstützt Sie beim Einlesen von Daten aus einer Nicht-SPSS-Datenquelle.
Hier können Sie die Datenquelle auswählen, angeben welche Fälle eingelesen werden sollen, die Dateien vor dem Einlesen aggregieren und sortieren und Variablennamen und -eigenschaften angeben. Einige Funktionen sind nur verfügbar, wenn eine Verbindung zum SPSS-Server besteht.
Aus welchen Datenquellen möchten Sie Daten lesen?
ODBC-Datenquellen
dBASE-Dateien
Excel-Dateien
Microsoft Access-Datenbank
SPSS
ODBC-Datenquelle hinzufügen...
< Zurück
Weiter >
Abbrechen
Hilfe

OLE DB-Datenquellen

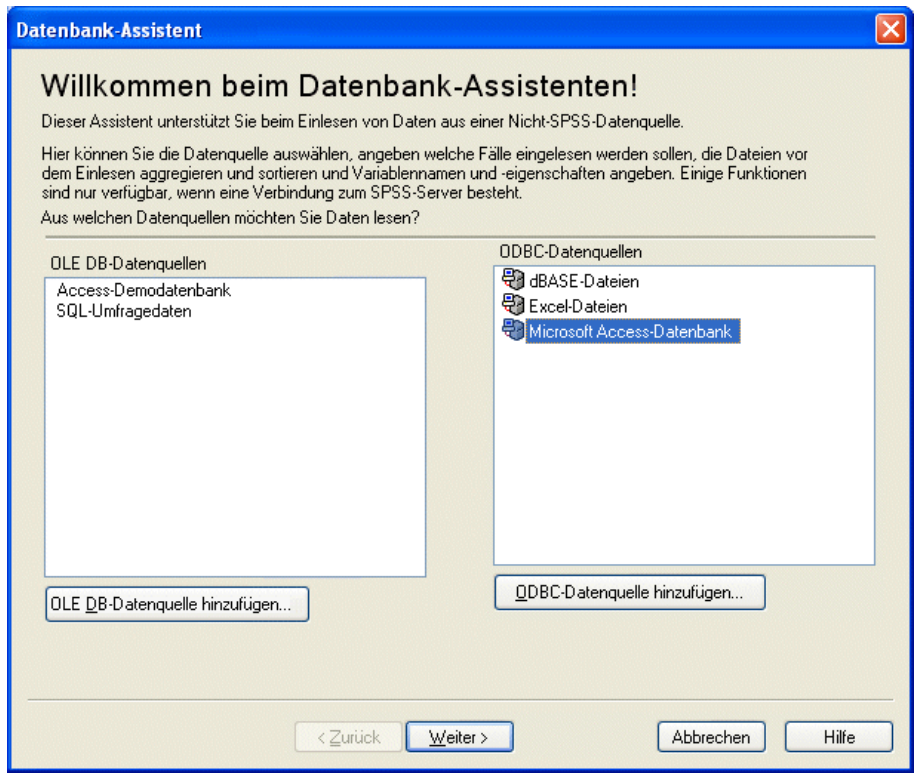
Wenn Sie auf OLE DB-Datenquellen zugreifen möchten, müssen folgende Komponenten auf dem Computer installiert sein, auf dem SPSS ausgeführt wird:

- .NET Framework
- Dimension-Datenmodell und OLE DB Access

Die Versionen dieser Komponenten, die mit dieser Version von SPSS kompatibel sind, können von der Installations-CD von SPSS installiert werden. Sie sind über das automatisch angezeigte Menü verfügbar.

- OLE DB-Datenquellen können nur im Modus für lokale Analysen hinzugefügt werden. Falls Sie OLE DB-Datenquellen auf einem Windows-Server im Modus für verteilte Analysen hinzufügen möchten, wenden Sie sich an Ihren Systemadministrator.
- Im Modus für verteilte Analysen (verfügbar mit SPSS Server) sind OLE DB-Datenquellen nur auf Windows-Servern verfügbar. Außerdem müssen sowohl .NET als auch das Dimensions-Datenmodell sowie OLE DB Access auf dem Server installiert sein.

Abbildung 3-2
Datenbank-Assistent mit Zugriff auf OLE DB-Datenquellen


Datenbank-Assistent
Willkommen beim Datenbank-Assistenten!
Dieser Assistent unterstützt Sie beim Einlesen von Daten aus einer Nicht-SPSS-Datenquelle.
Hier können Sie die Datenquelle auswählen, angeben welche Fälle eingelesen werden sollen, die Dateien vor dem Einlesen aggregieren und sortieren und Variablennamen und -eigenschaften angeben. Einige Funktionen sind nur verfügbar, wenn eine Verbindung zum SPSS-Server besteht.
Aus welchen Datenquellen möchten Sie Daten lesen?
OLE DB-Datenquellen
Access-Demodatenbank
SQL-Umfragedaten
OLE DB-Datenquelle hinzufügen...
ODBC-Datenquellen
dBASE-Dateien
Excel-Dateien
Microsoft Access-Datenbank
ODBC-Datenquelle hinzufügen...
< Zurück
Weiter >
Abbrechen
Hilfe

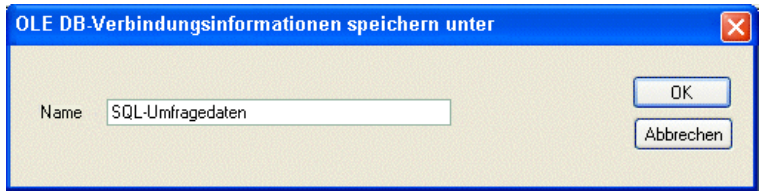
So fügen Sie eine OLE DB-Datenquelle hinzu:

- Klicken Sie auf OLE DB-Datenquelle hinzufügen.
- Klicken Sie im Dialogfeld “Eigenschaften der Datenverknüpfung” auf die Registerkarte Provider, und wählen Sie den OLE DB-Provider aus.
- Klicken Sie auf Weiter, oder klicken Sie auf die Registerkarte Verbindung.
- Wählen Sie die Datenbank aus, indem Sie das Verzeichnis und den Datenbanknamen eingeben, oder indem Sie auf die Schaltfläche klicken, um das Verzeichnis nach einer Datenbank zu durchsuchen. (Möglicherweise ist ein Benutzername und ein Passwort erforderlich.)

- Klicken Sie auf OK, nachdem Sie die erforderlichen Informationen eingegeben haben. (Durch Klicken auf die Schaltfläche Verbindung prüfen können Sie sicherstellen, dass die angegebene Datenbank verfügbar ist.)

- Geben Sie einen Namen für die Informationen zur Datenbankverbindung ein. (Dieser Name wird in der Liste der verfügbaren OLE DB-Datenquellen angezeigt.)

Abbildung 3-3
Dialogfeld "OLE DB-Verbindungsinformationen speichern unter"



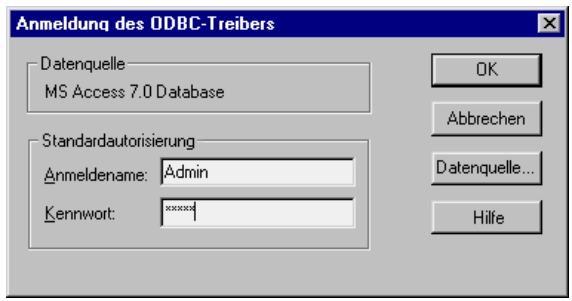
- Klicken Sie auf OK.

Hiermit gelangen Sie zurück zum ersten Bildschirm des Datenbank-Assistenten, auf dem Sie den gespeicherten Namen aus der Liste der OLE DB-Datenquellen auswählen und mit den weiteren Schritten des Assistenten fortfahren können.

Anmelden bei der Datenbank

Wenn für Ihre ODBC-Datenbank ein Passwort erforderlich ist, fordert Sie der Datenbank-Assistent zur Eingabe des Passworts auf, bevor die Datenquelle geöffnet werden kann.

Abbildung 3-4
Dialogfeld für das Anmelden bei der Datenbank



Auswählen von Datenfeldern

Im Dialogfeld “Daten auswählen” wird festgelegt, welche Tabellen und Felder eingelesen werden sollen. Datenbankfelder (Spalten) werden als Variablen in SPSS eingelesen.

Wenn in einer Tabelle eine beliebige Anzahl von Feldern ausgewählt wurde, werden im nächsten Fenster des Datenbank-Assistenten alle Felder dieser Tabelle angezeigt. Es werden jedoch nur die im Dialogfeld ausgewählten Felder als Variablen importiert. Auf diese Weise können Sie Verbindungen zwischen Tabellen erstellen und Kriterien festlegen, indem Sie die Felder verwenden, die nicht importiert werden.

Abbildung 3-5
Dialogfeld “Daten auswählen”

Anzeigen von Feldnamen. Klicken Sie auf das Pluszeichen (+) links neben dem Namen einer Tabelle, wenn Sie die Felder in dieser Tabelle einblenden möchten. Klicken Sie auf das Minuszeichen (–) links neben dem Namen einer Tabelle, wenn Sie die Felder ausblenden möchten.

So fügen Sie ein Feld hinzu. Doppelklicken Sie in der Liste “Verfügbare Tabellen” auf das entsprechende Feld, oder ziehen Sie dieses in das Listenfeld “Felder in dieser Reihenfolge einlesen”. Sie können die Reihenfolge der Felder ändern, indem Sie diese in der Liste der Felder ziehen und ablegen.

So entfernen Sie ein Feld. Doppelklicken Sie in der Liste “Felder in dieser Reihenfolge einlesen” auf das entsprechende Feld, oder ziehen Sie dieses in das Listenfeld “Verfügbare Tabellen”.

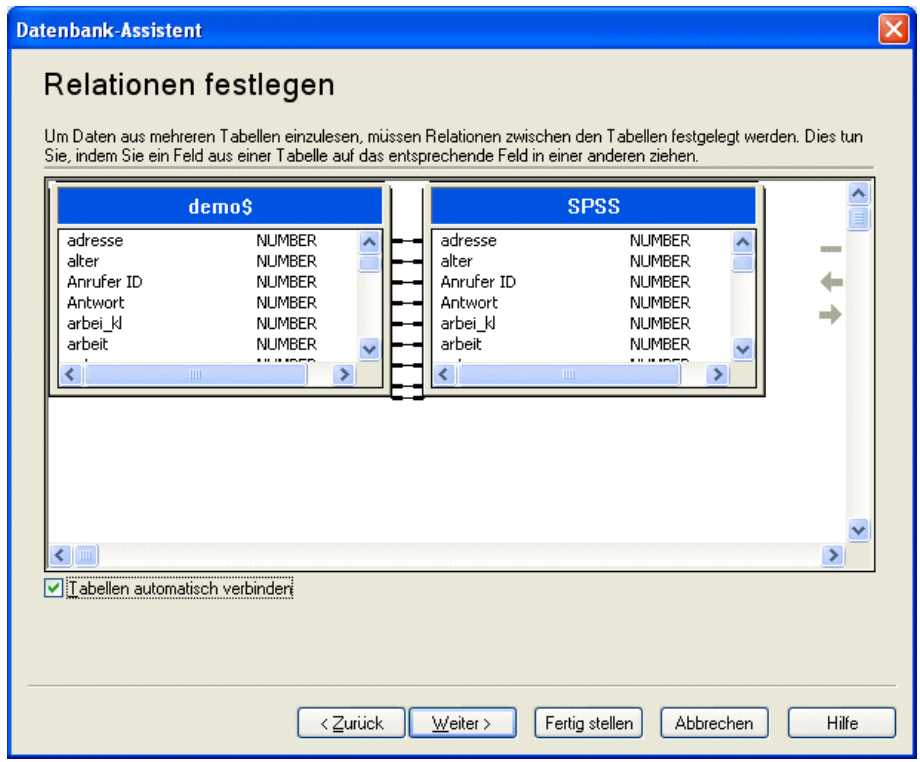
Feldnamen sortieren. Wenn dieses Kontrollkästchen aktiviert ist, werden die verfügbaren Felder im Datenbank-Assistenten in alphabetischer Reihenfolge angezeigt.

Anmerkung: Bei OLE DB-Datenquellen können Sie nur Felder aus einer einzelnen Tabelle auswählen. Verknüpfungen zwischen mehreren Tabellen werden bei OLE-DB-Datenquellen nicht unterstützt.

Erstellen einer Beziehung zwischen Tabellen

Im Dialogfeld “Relationen festlegen” können Sie für ODBC-Datenquellen die Relationen zwischen den Tabellen festlegen. Wenn Felder aus mehr als einer Tabelle ausgewählt sind, müssen Sie mindestens eine Verbindung festlegen.

Abbildung 3-6
Dialogfeld "Relationen festlegen"



Herstellen von Relationen. Zum Erstellen von Relationen ziehen Sie ein Feld aus einer beliebigen Tabelle auf das Feld, mit dem Sie dieses verbinden möchten. Im Datenbank-Assistenten wird dann eine **Verbindungslinie** zwischen den beiden Feldern angezeigt. Diese stellt die Beziehung dar. Die Datentypen der beiden Felder müssen übereinstimmen.

Tabellen automatisch verbinden. Hierbei wird versucht, Tabellen anhand von Primär-/Fremdschlüsseln oder übereinstimmenden Feldnamen und Datentypen automatisch zu verbinden.

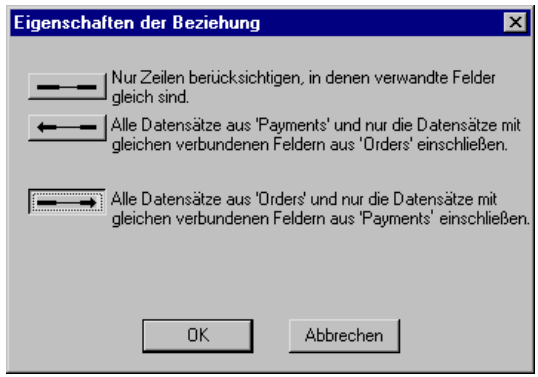
Festlegen der Verbindungstypen. Wenn der von Ihnen eingesetzte Treiber äußere Verbindungen unterstützt, können Sie innere, linke äußere und rechte äußere Verbindungen festlegen. Um den Typ der Verbindung auszuwählen, doppelklicken

Sie auf die Verbindungslinie zwischen den Feldern. Damit rufen Sie das Dialogfeld “Eigenschaften der Beziehung” auf.

Eigenschaften der Beziehung

In diesem Dialogfeld können Sie festlegen, durch welche Art von Beziehung Ihre Tabellen verbunden sind.

Abbildung 3-7
Dialogfeld “Eigenschaften der Beziehung”


Eigenschaften der Beziehung
Nur Zeilen berücksichtigen, in denen verwandte Felder gleich sind.
Alle Datensätze aus 'Payments' und nur die Datensätze mit gleichen verbundenen Feldern aus 'Orders' einschließen.
Alle Datensätze aus 'Orders' und nur die Datensätze mit gleichen verbundenen Feldern aus 'Payments' einschließen.
OK
Abbrechen

Innere Verknüpfungen. Eine innere Verknüpfung enthält nur die Zeilen, bei denen die verbundenen Felder übereinstimmen. In diesem Beispiel werden alle Zeilen mit übereinstimmenden ID-Werten in beiden Tabellen berücksichtigt.

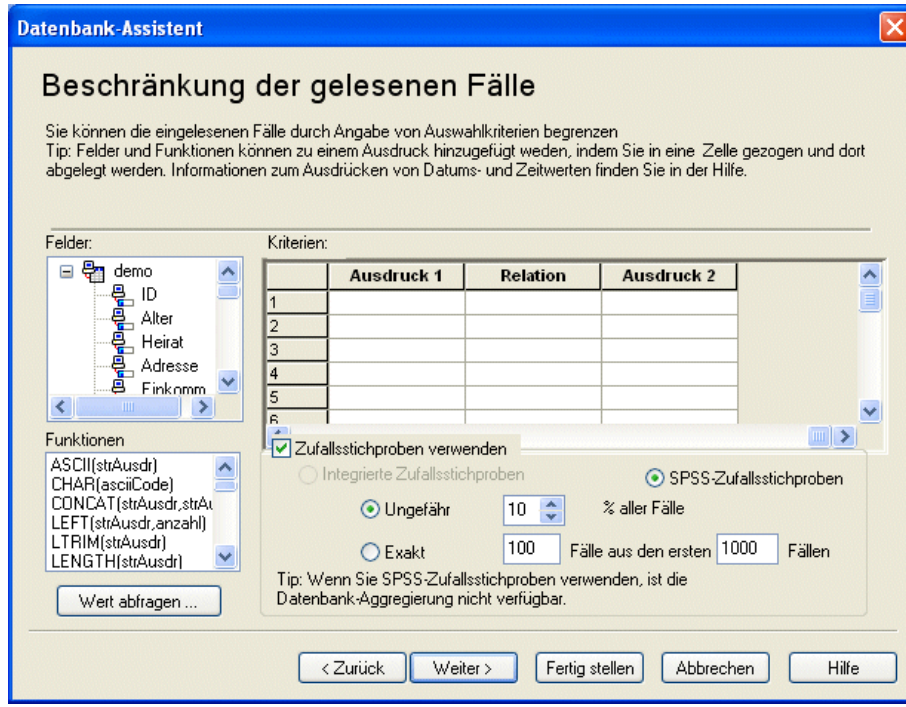
Äußere Verknüpfungen. Mit einer inneren Verknüpfung lassen sich Zuordnungen in Tabellen herstellen, die in einer 1:1-Beziehung stehen. Mit einer äußeren Verknüpfung sind zusätzlich Zuordnungen von Tabellen in einer 1:n-Beziehung möglich. So können Sie beispielsweise eine Tabelle mit nur wenigen Datensätzen, die Datenwerte und entsprechende beschreibende Labels darstellen, einer Tabelle mit hunderten oder tausenden Datensätzen zuordnen, die Teilnehmer an einer Umfrage darstellen. Eine linke äußere Verknüpfung enthält alle Datensätze aus der Tabelle auf der linken Seite und nur die Datensätze aus der Tabelle auf der rechten Seite, bei denen die verknüpften Felder übereinstimmen. In einer rechten äußeren Verbindung werden alle Datensätze aus der Tabelle auf der rechten Seite und nur die Datensätze aus der Tabelle auf der linken Seite importiert, bei denen die verknüpften Felder übereinstimmen.

Beschränkung der gelesenen Fälle

Im Dialogfeld “Beschränkung der gelesenen Fälle” können Sie Kriterien festlegen, mit denen Teilmengen von Fällen (Zeilen) ausgewählt werden. Im Allgemeinen werden zum Beschränken von Fällen die Kriterien in die Kriterientabelle eingegeben. Kriterien bestehen aus zwei Ausdrücken und einer zwischen diesen festgelegten Beziehung. Die Ausdrücke geben für jeden Fall die Werte *Wahr*, *Falsch* oder *Fehlend* zurück.

- Wenn als Ergebnis der Wert *Wahr* vorliegt, wird der Fall ausgewählt.
- Wenn als Ergebnis der Wert *Falsch* oder *Fehlend* vorliegt, wird der Fall nicht ausgewählt.
- Bei den meisten Kriterien wird mindestens einer der sechs Vergleichsoperatoren (<, >, <=, >=, =, <>) verwendet.
- Bedingte Ausdrücke können Feldnamen, Konstanten, arithmetische Operatoren, numerische und andere Funktionen sowie logische Variablen enthalten. Sie können Felder, die nicht importiert werden sollen, als Variablen verwenden.

Abbildung 3-8
Dialogfeld "Beschränkung der gelesenen Fälle"



Wenn Sie ein Kriterium erstellen möchten, benötigen Sie mindestens zwei Ausdrücke und eine zwischen den Ausdrücken festgelegte Beziehung, um diese Ausdrücke zu verbinden.

- Wenn Sie einen Ausdruck erstellen möchten, wählen Sie eine der folgenden Methoden aus:
  - Geben Sie in einer Ausdruckszelle Feldnamen, Konstanten, arithmetische Operatoren, numerische und andere Funktionen oder logische Variablen ein.
  - Doppelklicken Sie in der Liste "Felder" auf das Feld.
  - Ziehen Sie das Feld aus der Liste "Felder" auf eine Ausdruckszelle.
  - Wählen Sie aus dem Dropdown-Menü einer beliebigen aktiven Ausdruckszelle ein Feld aus.

- Um einen relationalen Operator (z. B. = oder >) auszuwählen, setzen Sie die Einfügemarke in die Beziehungszelle, und geben Sie entweder den Operator manuell ein, oder wählen Sie diesen im Dropdown-Menü aus.

  Bei Daten und Uhrzeiten in Ausdrücken ist ein spezielles Format erforderlich (mit den geschweiften Klammern in den Beispielen):

  - Für Datumsangaben gilt die folgende allgemeine Form: {d yyyy-mm-dd}.
  - Für Uhrzeitangaben gilt die folgende allgemeine Form: {t hh:mm:ss}.
  - Für Datums- und Uhrzeitangaben gilt die folgende allgemeine Form: {dt yyyy-mm-dd hh:mm:ss}.

  Funktionen. SPSS stellt eine Reihe von arithmetischen und logischen SQL-Funktionen sowie SQL-Funktionen für Zeichenfolgen, Datumsangaben und Zeitangaben zur Verfügung. Sie können diese Funktionen aus der Liste auswählen und in den Ausdruck ziehen oder beliebige gültige SQL-Funktionen eingeben. Informationen zu den gültigen SQL-Funktionen finden Sie in der Dokumentation Ihrer Datenbank. Eine Liste der Standardfunktionen finden Sie unter:

  http://msdn.microsoft.com/library/en-us/odbc/htm/odbcscalar_functions.asp

  Zufallsstichproben verwenden. Mit dieser Option wird aus der Datenquelle eine Zufallsstichprobe von Fällen ausgewählt. Bei großen Datenquellen soll die Anzahl der Fälle möglicherweise auf eine kleine, repräsentative Auswahl begrenzt werden, womit die Laufzeit von Prozeduren beträchtlich verringert werden kann. Integrierte Zufallsstichproben sind, falls für die Datenquelle verfügbar, schneller als SPSS-Zufallsstichproben, da bei letzteren noch die gesamte Datenquelle gelesen werden muss, um eine Zufallsstichprobe zu extrahieren.

  - **Ungefähr (Select Random Sample).** Erzeugt eine Zufallsstichprobe, die näherungsweise den angegebenen Prozentsatz der Fälle enthält.
  - **Exakt (Select Random Sample).** Wählt eine Zufallsstichprobe mit der angegebenen Anzahl von Fällen aus der festgelegten Gesamtanzahl der Fälle aus. Wenn die angegebene Gesamtanzahl der Fälle größer als die Anzahl der Fälle in der Datei ist, enthält die Zufallsstichprobe anteilig weniger Fälle, als angegeben wurde.

  Anmerkung: Bei Zufallsstichproben steht die Aggregation (verfügbar im Modus für verteilte Analysen mit SPSS Server) nicht zur Verfügung.

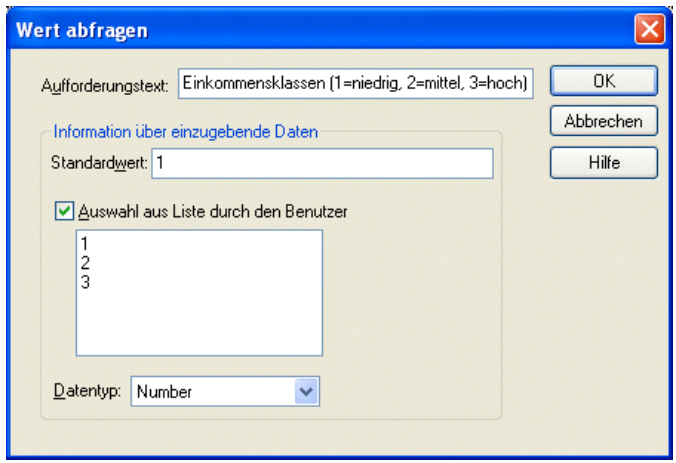
Wert abfragen. Zum Erstellen einer **Parameterabfrage** können Sie in die Abfrage eine Eingabeaufforderung integrieren. Bei Ausführen der Abfrage werden die Benutzer dann aufgefordert, anhand der dieser Angaben Informationen einzugeben. Eingabeaufforderungen können in Situationen nützlich sein, in denen verschiedene Ansichten derselben Daten benötigt werden. Sie möchten beispielsweise die Verkaufszahlen für verschiedene Rechnungsjahre unter Verwendung derselben Abfrage einsehen.

- Setzen Sie die Einfügemarke in eine beliebige Ausdruckszelle, und klicken Sie zum Erstellen einer Eingabeaufforderung auf Wert abfragen.

Erstellen einer Parameterabfrage

Mit dem Dialogfeld "Wert abfragen" können Sie ein Dialogfeld erstellen, in dem die Benutzer bei jedem neuen Ausführen einer Abfrage aufgefordert werden, die entsprechenden Informationen einzugeben. Diese Funktion ist nützlich, wenn dieselbe Datenquelle anhand von verschiedenen Kriterien abgefragt werden soll.

Abbildung 3-9
Dialogfeld "Wert abfragen"



Geben Sie zum Erstellen einer Eingabeaufforderung den Aufforderungstext und einen Standardwert ein. Der Aufforderungstext wird immer dann angezeigt, wenn ein Benutzer die Abfrage ausführt. Der Text sollte die Art der einzugebenden Information beschreiben. Wenn der Benutzer nicht aus einer Liste auswählt, sollte der Text einen

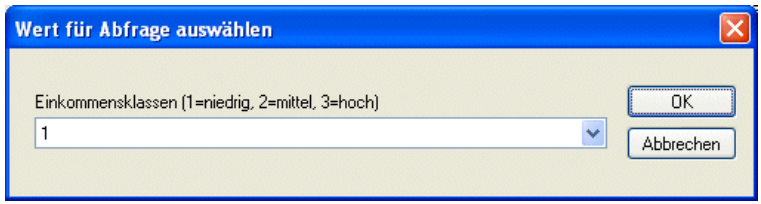
Hinweis darauf geben, wie die Eingabe formatiert werden soll. Ein Beispiel lautet folgendermaßen: Geben Sie ein Quartal ein (Q1, Q2, Q3, ...).

Auswahl aus Liste durch den Benutzer. Wenn dieses Kontrollkästchen aktiviert ist, wird der Benutzer auf die Auswahl der von Ihnen bereitgestellten Werte beschränkt. Stellen Sie sicher, dass die Werte durch Zeilenumbrüche getrennt sind.

Datentyp. Wählen Sie hier den zu verwendenden Datentyp aus (Numerisch, String oder Datum).

Das abschließende Ergebnis sieht folgendermaßen aus:

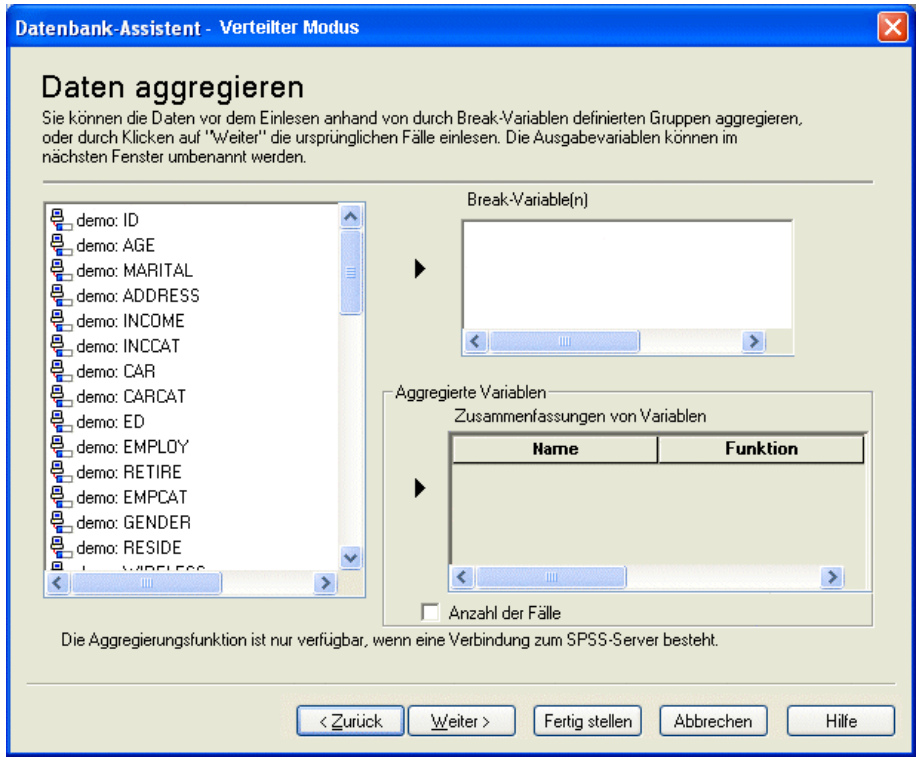
Abbildung 3-10
Dialogfeld für benutzerdefinierte Eingabeaufforderung



Aggregieren von Daten

Wenn Sie sich im Modus für verteilte Analysen befinden und eine Verbindung zu einem Remote-Server besteht (mit SPSS Server verfügbar), können Sie die Daten vor dem Einlesen in SPSS aggregieren.

Abbildung 3-11
Dialogfeld "Daten aggregieren"



Die Daten können auch nach dem Einlesen in SPSS aggregiert werden; bei umfangreichen Datenquellen kann die vorherige Aggregierung jedoch beträchtlich Zeit sparen.

- Um aggregierte Daten zu erstellen, wählen Sie eine oder mehrere Break-Variablen aus, die definieren, wie die Fälle gruppiert werden.
- Wählen Sie mindestens eine aggregierte Variable aus.
- Wählen Sie für jede Aggregierungsvariable eine Aggregierungsfunktion aus.
- Sie können bei Bedarf auch eine Variable anlegen, welche die Anzahl der Fälle in jeder Break-Gruppe enthält.

  Anmerkung: Bei Zufallsstichproben ist die Aggregation nicht verfügbar.

Definieren von Variablen

Variablennamen und -labels. In SPSS wird der vollständige Name des Datenbankfelds (der Spalte) als Variablenlabel verwendet. Wenn Sie keine Änderungen an den Variablennamen vornehmen, weist der Datenbank-Assistent jeder Spalte der Datenbank selbständig einen Variablennamen zu. Bei der Vergabe von Variablennamen werden die beiden folgenden Verfahren eingesetzt:

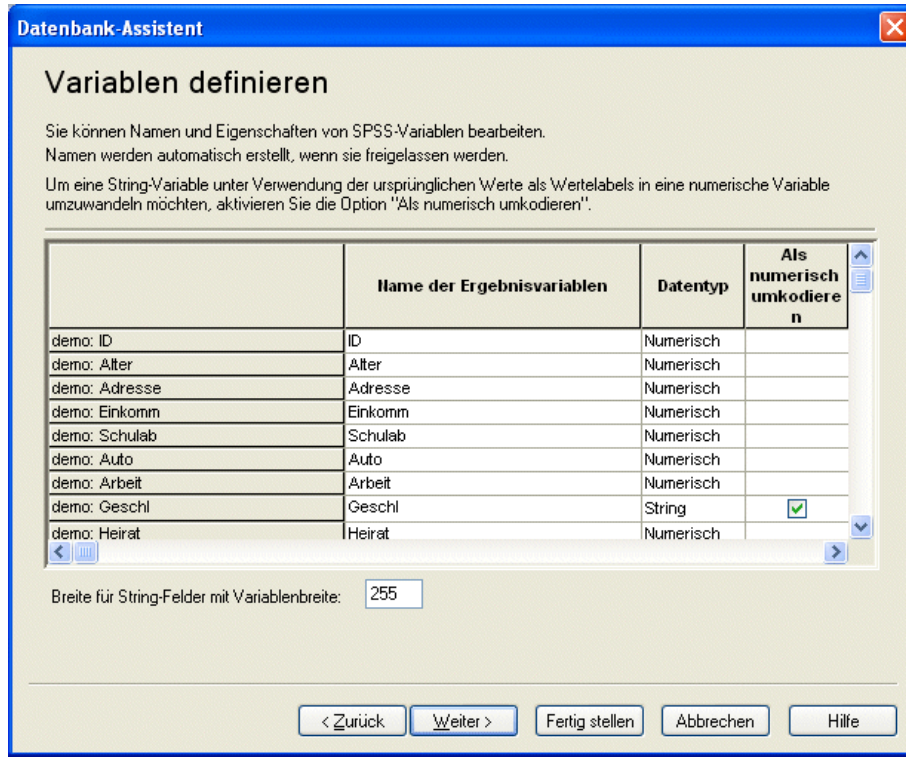
- Wenn der Name des Datenbankfelds einen gültigen und eindeutigen Variablennamen für SPSS ergibt, wird dieser Name als Variablenname verwendet.
- Wenn der Name des Datenbankfelds keinen gültigen und eindeutigen Variablennamen ergibt, wird automatisch ein neuer, eindeutiger Name erstellt.

Klicken Sie auf eine beliebige Zelle, um den Variablennamen zu bearbeiten.

Umwandeln von Strings in numerische Werte. Wählen Sie das Feld Als numerisch umkodieren bei einer String-Variablen aus, wenn diese automatisch in eine numerische Variable umgewandelt werden soll. String-Werte werden anhand der alphabetischen Reihenfolge der ursprünglichen Werte in fortlaufende, ganzzahlige Werte umgewandelt. Die ursprünglichen Werte werden als Wertelabels für die neuen Variablen beibehalten.

Breite für String-Felder mit Variablenbreite. Mit dieser Option wird die Breite der String-Werte mit Variablenbreite gesteuert. Standardmäßig beträgt die Breite 255 Byte, und nur die ersten 255 Byte (in der Regel 255 Zeichen bei Single-Byte-Sprachen) werden gelesen. Die Breite kann bis zu 32.767 Byte umfassen. In der Regel sollen String-Werte zwar nicht gekürzt werden, aber auch übermäßig große Werte sollten vermieden werden, weil dies zu Leistungseinbußen bei der Verarbeitung mit SPSS führt.

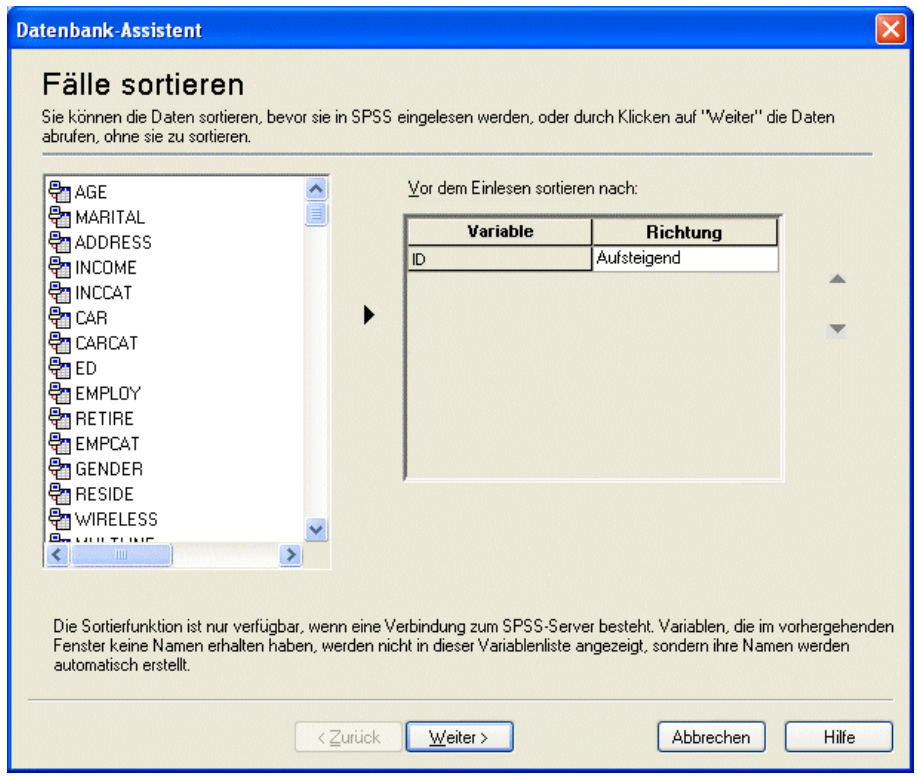
Abbildung 3-12
Dialogfeld "Variablen definieren"



Sortieren von Fällen

Wenn Sie sich im Modus für verteilte Analysen befinden und eine Verbindung zu einem Remote-Server besteht (mit SPSS Server verfügbar), können Sie die Daten vor dem Einlesen in SPSS sortieren.

Abbildung 3-13
Dialogfeld "Fälle sortieren"



Die Daten können auch nach dem Einlesen in SPSS sortiert werden; bei umfangreichen Datenquellen kann die vorherige Sortierung jedoch beträchtlich Zeit sparen.

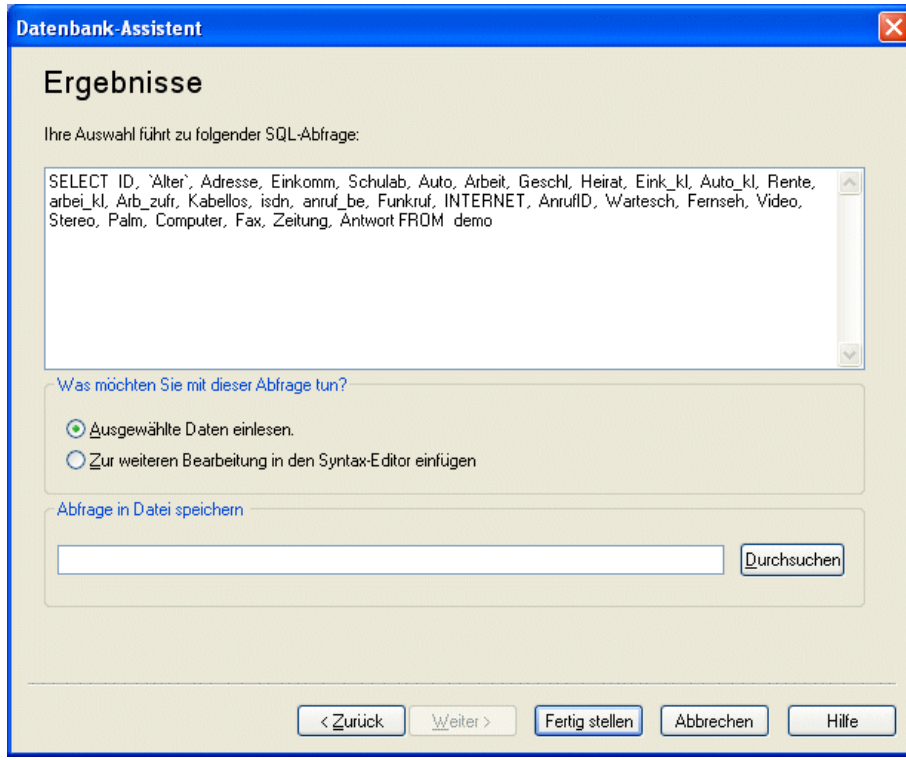
Ergebnisse

Im Dialogfeld "Ergebnisse" wird die SQL-SELECT-Anweisung der Abfrage angezeigt.

- Sie können die SQL-SELECT-Anweisung vor dem Ausführen der Abfrage bearbeiten. Wenn Sie jedoch auf die Schaltfläche Zurück klicken, um Änderungen in vorangegangenen Schritten vorzunehmen, gehen die Änderungen an der SELECT-Anweisung verloren.
- Verwenden Sie den Abschnitt Abfrage in Datei speichern, um die Abfrage für eine zukünftige Verwendung zu speichern.
- Um die vollständige Syntax für `GET DATA` in ein Syntax-Fenster einzufügen, wählen Sie Zur weiteren Bearbeitung in den Syntax-Editor einfügen aus. Durch einfaches Kopieren und Einfügen der SELECT-Anweisung aus dem Ergebnisfeld wird nicht die erforderliche Befehlssyntax eingefügt.

Anmerkung: Die eingefügte Syntax enthält in jeder SQL-Zeile, die vom Assistenten erzeugt wurde, ein Leerzeichen vor dem schließenden Anführungszeichen. Diese Leerzeichen sind nicht überflüssig. Bei der Verarbeitung des Befehls werden alle Zeilen in der SQL-Anweisung zusammengeführt. Ohne das Leerzeichen würde das letzte Zeichen in einer Zeile übergangslos am ersten Zeichen in der jeweils nachfolgenden Zeile “kleben”.

Abbildung 3-14
Dialogfeld "Ergebnisse"


Datenbank-Assistent
Ergebnisse
Ihre Auswahl führt zu folgender SQL-Abfrage:
SELECT ID, `Alter`, Adresse, Einkomm, Schulab, Auto, Arbeit, Geschl, Heirat, Eink_kl, Auto_kl, Rente, arbei_kl, Arb_zufr, Kabellos, isdn, anruf_be, Funkruf, INTERNET, AnrufID, Wartesch, Fernseh, Video, Stereo, Palm, Computer, Fax, Zeitung, Antwort FROM demo
Was möchten Sie mit dieser Abfrage tun?
Ausgewählte Daten einlesen.
Zur weiteren Bearbeitung in den Syntax-Editor einfügen
Abfrage in Datei speichern
Durchsuchen
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

Text-Assistent

Der Text-Assistent kann Textdatendateien in einer Vielzahl von verschiedenen Formaten einlesen:

- Tabulator-getrennte Dateien
- Leerzeichen-getrennte Dateien
- Komma-getrennte Dateien
- Dateien mit Spalten fester Breite

Sie können bei Dateien mit Trennzeichen auch andere Zeichen als Trennzeichen zwischen Werten festlegen, und Sie können mehrere Trennzeichen bestimmen.

So lesen Sie Textdatendateien ein

- Wählen Sie die folgenden Befehle aus den Menüs aus:

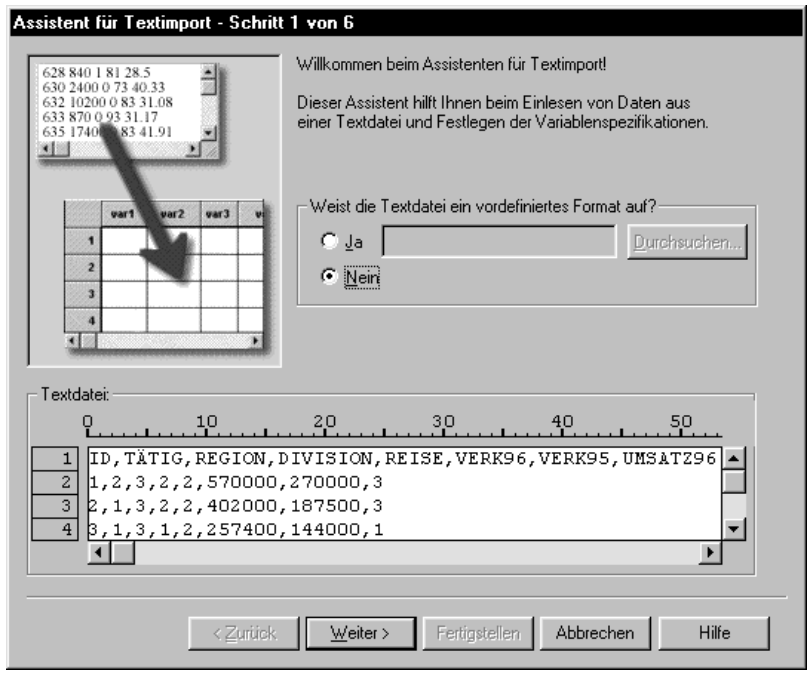
  Datei
  Textdaten einlesen...

- Wählen Sie im Dialogfeld “Datei öffnen” die Textdatei aus.

- Bestimmen Sie, wie die Datendatei eingelesen werden soll. Gehen Sie dazu anhand der Schritte im Text-Assistenten vor.

Text-Assistent: Schritt 1

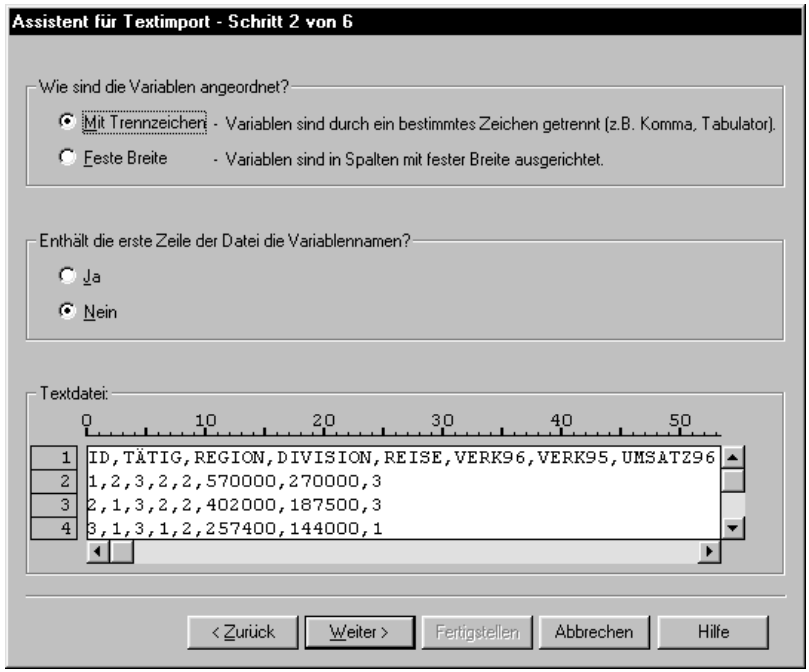
Abbildung 3-15
Text-Assistent: Schritt 1



Die Textdatei wird im Vorschaufenster angezeigt. Sie können ein vordefiniertes (also zuvor im Text-Assistenten gespeichertes) Format anwenden oder anhand der Schritte im Text-Assistenten vorgehen, um festzulegen, wie die Daten eingelesen werden sollen.

Text-Assistent: Schritt 2

Abbildung 3-16
Text-Assistent: Schritt 2


Assistent für Textimport - Schritt 2 von 6
Wie sind die Variablen angeordnet?
Mit Trennzeichen - Variablen sind durch ein bestimmtes Zeichen getrennt (z.B. Komma, Tabulator).
Feste Breite - Variablen sind in Spalten mit fester Breite ausgerichtet.
Enthält die erste Zeile der Datei die Variablennamen?
Ja
Nein
Textdatei:
0 10 20 30 40 50
1 ID,TÄTIG,REGION,DIVISION,REISE,VERK96,VERK95,UMSATZ96
2 1,2,3,2,2,570000,270000,3
3 2,1,3,2,2,402000,187500,3
4 3,1,3,1,2,257400,144000,1
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

In diesem Schritt erhalten Sie Informationen über die Variablen. Eine Variable ist vergleichbar mit einem Feld in einer Datenbank. So stellt beispielsweise jede Frage auf einem Fragebogen eine Variable dar.

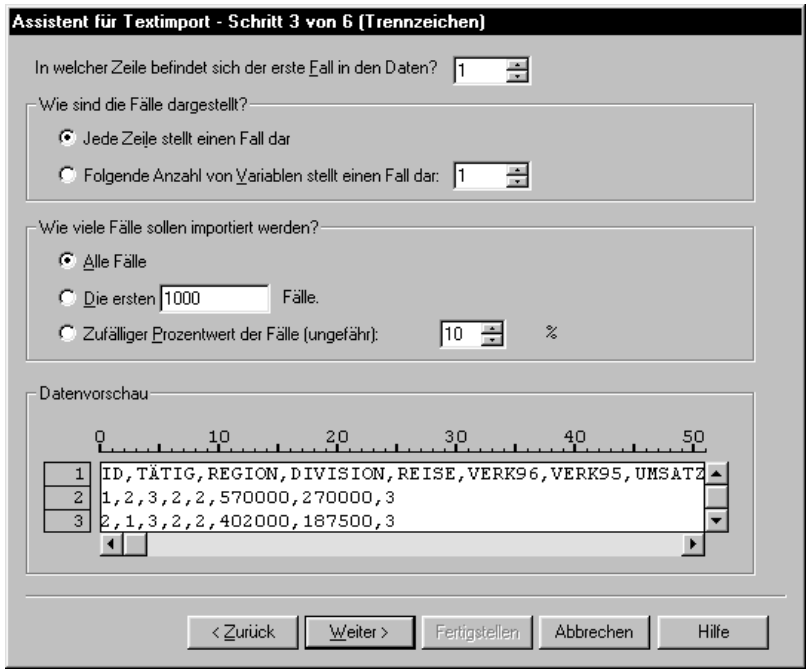
Wie sind die Variablen angeordnet? Damit der Text-Assistent die Daten ordnungsgemäß einlesen kann, müssen Sie dem Text-Assistenten mitteilen, woran das Ende des Datenwerts einer Variablen bzw. der Anfang des Datenwerts der nachfolgenden Variablen erkannt werden kann. Die Methode, mit der eine Variable von der nächsten unterschieden wird, ist abhängig von der Anordnung der Variablen.

- **Mit Trennzeichen.** Leerzeichen, Kommata, Tabulatoren oder andere Zeichen werden zum Trennen von Variablen verwendet. Die Variablen werden für jeden Fall in derselben Reihenfolge aufgezeichnet, befinden sich aber nicht notwendigerweise an derselben Spaltenposition.
- **Feste Breite.** Jede Variable wird für jeden Fall in der Datendatei an der gleichen Spaltenposition in derselben Zeile im Datensatz aufgezeichnet. Zwischen den Variablen sind keine Trennzeichen erforderlich. Bei vielen durch Computerprogramme erzeugten Textdatendateien kann es den Anschein haben, daß die Datenwerte ohne Trennzeichen ineinanderlaufen. Durch die Spaltenposition ist jedoch festgelegt, welche Variable eingelesen wird.

Enthält die erste Zeile der Datei die Variablennamen? Wenn die erste Zeile der Datendatei aussagekräftige Beschreibungen für jede Variable enthält, können Sie diese Beschreibungen als Variablennamen einsetzen. Werte, die nicht den Regeln für Variablennamen entsprechen, werden in gültige Variablennamen umgewandelt.

Text-Assistent: Schritt 3 (Durch Trennzeichen getrennte Dateien)

Abbildung 3-17
Text-Assistent: Schritt 3 (für durch Trennzeichen getrennte Dateien)



In diesem Schritt erhalten Sie Informationen über die Fälle. Ein Fall ist vergleichbar mit einem Datensatz in einer Datenbank. So ist zum Beispiel jede Person, die einen Fragebogen ausfüllt, ein Fall.

Auf welcher Zeile befindet sich der erste Fall in den Daten? Gibt die erste Zeile der Datendatei an, die Datenwerte enthält. Wenn die erste(n) Zeile(n) der Datendatei aussagekräftige Beschreibungen oder anderen Text und somit keine Datenwerte enthalten, ist dies *nicht* Zeile 1.

Wie sind die Fälle dargestellt? Gibt an, woran der Text-Assistent erkennt, wo jeder Fall endet und der nächste beginnt.

- **Jede Zeile stellt einen Fall dar.** Jede Zeile enthält nur einen Fall. Sehr häufig ist jeder Fall in einer einzigen Zeile enthalten. Hierbei kann es sich auch um eine sehr lange Zeile in einer Datendatei mit einer großen Anzahl von Variablen handeln. Wenn nicht alle Zeilen die gleiche Anzahl von Werten enthalten,

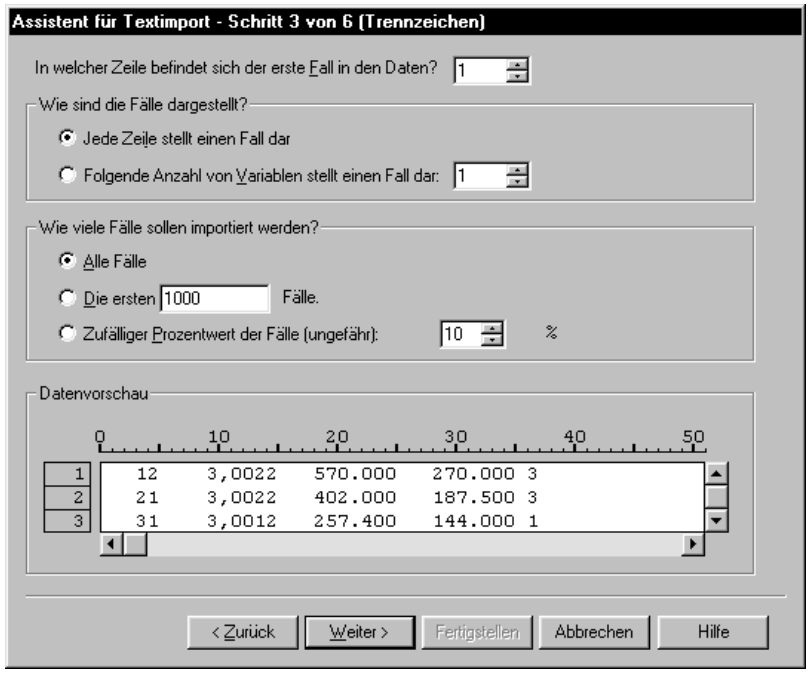
wird die Anzahl der Variablen für jeden Fall durch die Zeile mit den meisten Datenwerten festgelegt. Fällen mit weniger Datenwerten werden fehlende Werte für die zusätzlichen Variablen zugewiesen.

- **Folgende Anzahl von Variablen stellt einen Fall dar.** Durch die festgelegte Anzahl von Variablen pro Fall wird dem Text-Assistenten mitgeteilt, an welcher Stelle eine Variable endet und die nächste beginnt. Eine Zeile kann mehrere Fälle enthalten, und Fälle können in der Mitte einer Zeile beginnen und in der nächsten Zeile fortgesetzt werden. Der Text-Assistent bestimmt das Ende jedes Falls unabhängig von der Anzahl der Zeilen anhand der Anzahl von eingelesenen Werten. Für jeden Fall müssen Datenwerte (oder durch Trennzeichen angezeigte fehlende Werte) für alle Variablen vorhanden sein. Ansonsten wird die Datendatei nicht ordnungsgemäß eingelesen.

Wie viele Fälle sollen importiert werden? Sie können alle Fälle der Datendatei, die ersten *n* Fälle (*n* ist eine von Ihnen festgelegte Zahl) oder eine Stichprobe mit einem bestimmten Prozentsatz der Fälle importieren. Da SPSS für jeden Fall eine unabhängige Pseudo-Zufallsentscheidung trifft, entspricht der Prozentsatz der tatsächlich ausgewählten Fälle nur ungefähr dem angegebenen Prozentwert. Je mehr Fälle sich in der Datendatei befinden, desto eher entspricht der Prozentsatz ausgewählter Fälle dem angegebenen Prozentsatz.

Text-Assistent: Schritt 3 (Dateien mit Spalten fester Breite)

Abbildung 3-18
Text-Assistent: Schritt 3 (für Dateien mit Spalten fester Breite)



In diesem Schritt erhalten Sie Informationen über die Fälle. Ein Fall ist vergleichbar mit einem Datensatz in einer Datenbank. So ist zum Beispiel jede Person, die einen Fragebogen ausfüllt, ein Fall.

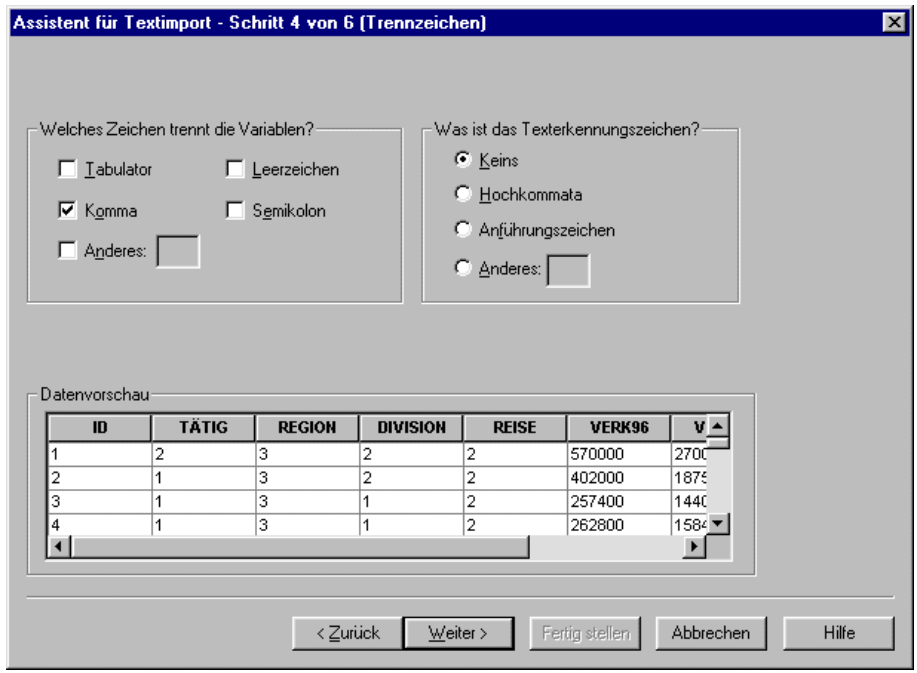
Auf welcher Zeile befindet sich der erste Fall in den Daten? Gibt die erste Zeile der Datendatei an, die Datenwerte enthält. Wenn die erste(n) Zeile(n) der Datendatei aussagekräftige Beschreibungen oder anderen Text und somit keine Datenwerte enthalten, ist dies *nicht* Zeile 1.

Wie viele Zeilen stellen einen Fall dar? Gibt an, woran der Text-Assistent erkennt, wo jeder Fall endet und der nächste beginnt. Jede Variable ist durch ihre Zeilennummer innerhalb des Falls und durch ihre Spaltenposition festgelegt. Sie müssen die Anzahl der Zeilen pro Fall angeben, damit die Daten richtig eingelesen werden können.

Wie viele Fälle sollen importiert werden? Sie können alle Fälle der Datendatei, die ersten *n* Fälle (*n* ist eine von Ihnen festgelegte Zahl) oder eine Stichprobe mit einem bestimmten Prozentsatz der Fälle importieren. Da SPSS für jeden Fall eine unabhängige Pseudo-Zufallsentscheidung trifft, entspricht der Prozentsatz der tatsächlich ausgewählten Fälle nur ungefähr dem angegebenen Prozentwert. Je mehr Fälle sich in der Datendatei befinden, desto eher entspricht der Prozentsatz ausgewählter Fälle dem angegebenen Prozentsatz.

Text-Assistent: Schritt 4 (Durch Trennzeichen getrennte Dateien)

Abbildung 3-19
Text-Assistent: Schritt 4 (für durch Trennzeichen getrennte Dateien)



In diesem Schritt zeigt der Text-Assistent einen Vorschlag an, wie die Variablen aus der Datendatei eingelesen werden. Sie können diesen Vorschlag ändern.

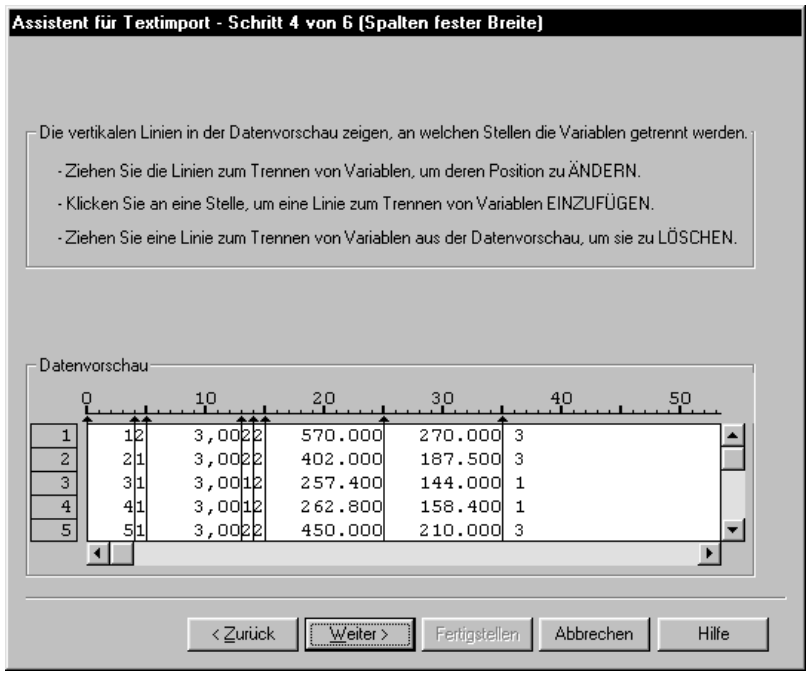
Welches Zeichen trennt die Variablen? Geben Sie hier das Zeichen an, mit dem die Datenwerte getrennt werden. Sie können eine beliebige Kombination aus Leerzeichen, Kommata, Semikola, Tabulatoren und anderen Zeichen wählen.

Mehrere aufeinanderfolgende Trennzeichen, die keine Datenwert einschließen, werden als fehlende Werte behandelt.

Was ist ein Texterkennungszeichen? Hierbei handelt es sich um Zeichen, mit denen Werte eingeschlossen werden, die Trennzeichen enthalten. Wenn beispielsweise ein Komma das Trennzeichen ist, werden Werte mit Kommata falsch eingelesen, sofern es keine Texterkennungszeichen gibt, die den Wert einschließen und damit verhindern, daß die Kommata in dem Wert als Trennzeichen zwischen Werten interpretiert werden. Bei Datendateien im CSV-Format, die aus Excel exportiert werden, werden Anführungszeichen (“) als Texterkennungszeichen verwendet. Die Texterkennungszeichen erscheinen am Anfang und am Ende des Werts, umschließen also den ganzen Wert.

Text-Assistent: Schritt 4 (Dateien mit Spalten fester Breite)

Abbildung 3-20
Text-Assistent: Schritt 4 (für Dateien mit Spalten fester Breite)


Assistent für Textimport - Schritt 4 von 6 (Spalten fester Breite)
Die vertikalen Linien in der Datenvorschau zeigen, an welchen Stellen die Variablen getrennt werden.
- Ziehen Sie die Linien zum Trennen von Variablen, um deren Position zu ÄNDERN.
- Klicken Sie an eine Stelle, um eine Linie zum Trennen von Variablen EINZUFÜGEN.
- Ziehen Sie eine Linie zum Trennen von Variablen aus der Datenvorschau, um sie zu LÖSCHEN.
Datenvorschau
0
10
20
30
40
50
1 12 3,00 22 570.000 270.000 3
2 21 3,00 22 402.000 187.500 3
3 31 3,00 12 257.400 144.000 1
4 41 3,00 12 262.800 158.400 1
5 51 3,00 22 450.000 210.000 3
< Zurück
Weiter >
Fertigstellen
Abbrechen
Hilfe

In diesem Schritt zeigt der Text-Assistent einen Vorschlag an, wie die Variablen aus der Datendatei eingelesen werden. Sie können diesen Vorschlag ändern. Vertikale Linien im Vorschaufenster kennzeichnen die Positionen, von denen der Text-Assistent annimmt, daß sie den Anfang der einzelnen Variablen in der Datei kennzeichnen.

Fügen Sie erforderlichenfalls weitere Variablentrennlinien ein, oder verschieben bzw. löschen Sie vorhandene Linien. Wenn für jeden Fall mehrere Zeilen verwendet werden, wählen Sie die einzelnen Zeilen aus der Dropdown-Liste und nehmen die notwendigen Änderungen an den Variablentrennlinien vor.

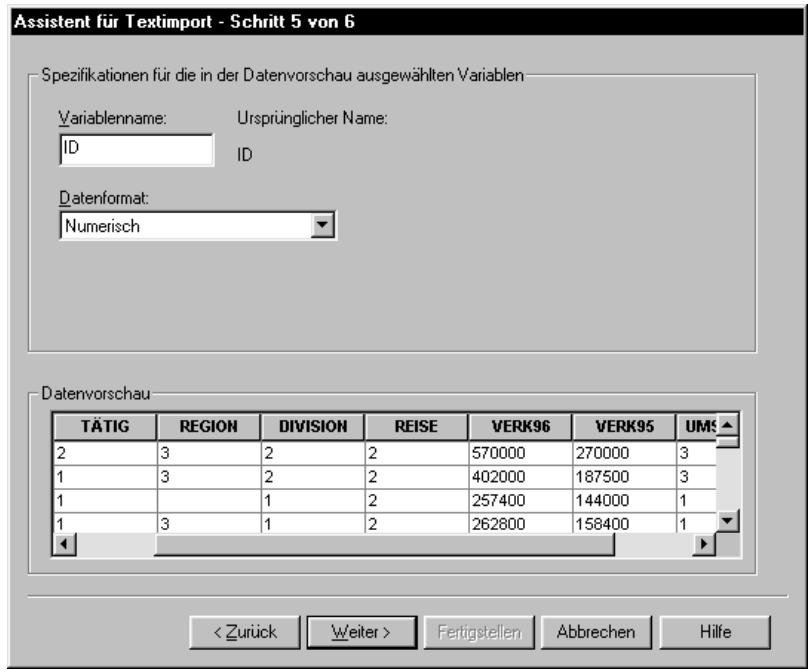
Anmerkungen:

Bei computergenerierten Datendateien mit einem kontinuierlichen Fluß an Datenwerten ohne trennende Leerzeichen oder andere Zeichen kann es große Schwierigkeiten bereiten, zu bestimmen, wo die einzelnen Variablen beginnen. Bei dieser Art von Datendatei sind normalerweise Datendefinitionsdateien oder andere schriftliche Beschreibungen notwendig, in denen die Zeile und die Spalte für jede Variable aufgeführt sind.

Zum Anzeigen des Dateiinhalts im Text-Assistenten sollte eine Schriftart mit fester Zeichenbreite verwendet werden. Bei Schriftarten mit variabler Zeichenbreite wird der Dateiinhalt nicht ordnungsgemäß ausgerichtet. Die Schriftart wird im Feld “Schriftart für Textausgabe” auf der Registerkarte “Viewer” des Dialogfelds “Optionen” (Menü “Bearbeiten”) festgelegt.

Text-Assistent: Schritt 5

Abbildung 3-21
Text-Assistent: Schritt 5



In diesem Schritt legen Sie fest, welchen Variablennamen und welches Datenformat der Text-Assistent beim Einlesen der Variablen verwendet und welche Variablen in der endgültigen Datendatei enthalten sein werden.

Variablenname. Sie können die vom System vorgegebenen Variablennamen durch Ihre eigenen überschreiben. Beim Einlesen von Variablennamen aus der Datendatei ändert der Text-Assistent automatisch Variablennamen, die nicht den Regeln für Variablennamen entsprechen. Wählen Sie im Vorschaufenster eine Variable, und geben Sie einen Variablennamen ein.

Datenformat. Wählen Sie im Vorschaufenster eine Variable, und wählen Sie ein Format aus der Dropdown-Liste aus. Halten Sie zum Auswählen mehrerer aufeinanderfolgender Variablen beim Klicken die UMSCHALTTASTE gedrückt. Halten Sie beim Auswählen von mehreren nicht aufeinanderfolgenden Variablen beim Klicken die STRG-Taste gedrückt.

Formatoptionen im Text-Assistenten

Beim Einlesen von Variablen mit dem Text-Assistenten sind die folgenden Formatoptionen verfügbar:

Nicht importieren. Die gewählte(n) Variable(n) in der Datendatei wird/werden übersprungen.

Numerisch. Gültige Werte sind Ziffern, ein führendes Plus- oder Minuszeichen und ein Dezimaltrennzeichen.

String. Gültige Werte hierfür sind fast alle Zeichen, die auf der Tastatur eingegeben werden können, und eingebettete Leerzeichen. Bei Dateien mit Trennzeichen können Sie die Anzahl der Zeichen im Wert angeben, bis zu maximal 32,767. In der Standardeinstellung setzt der Text-Assistent die Zeichenanzahl auf den längsten String, der für die gewählte(n) Variable(n) gefunden werden kann. Bei Dateien mit Spalten fester Breite wird die Zeichenanzahl der Strings durch die Anordnung der Variablentrennlinien in Schritt 4 bestimmt.

Datum/Uhrzeit. Zu den gültigen Werten zählen Daten im allgemeinen Format *tt-mm-jjjj*, *mm/tt/jjjj*, *tt.mm.jjjj*, *jjjj/mm/tt*, *hh:mm:ss* und eine Vielzahl anderer Formate für das Datum und die Uhrzeit. Monate können durch arabische oder römische Ziffern und dreibuchstabige Abkürzungen dargestellt oder vollständig ausgeschrieben werden. Wählen Sie ein Datumsformat aus der Liste aus.

Dollar. Als zulässige Werte gelten Ziffern mit optionalem führenden Dollarzeichen und optionalen Kommata als Tausendertrennzeichen.

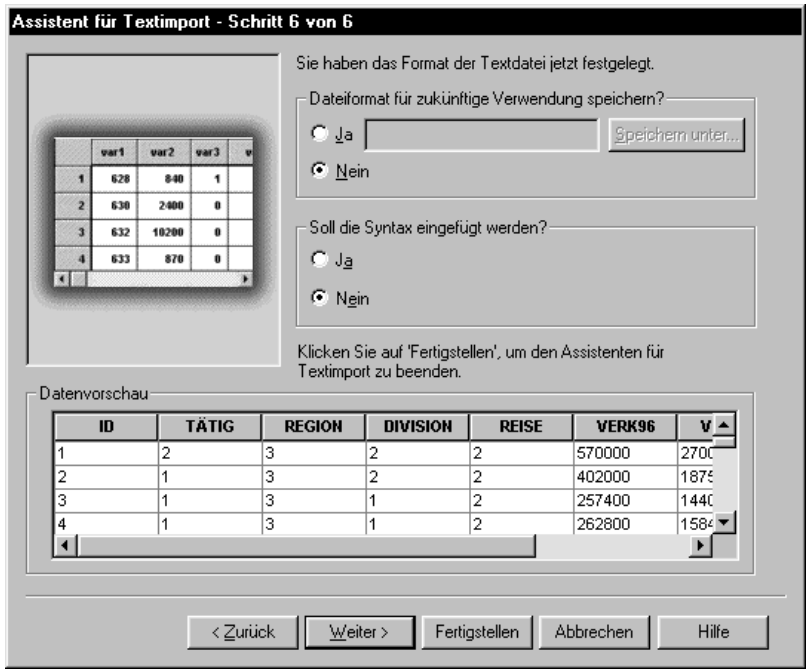
Komma. Als gültige Werte hierfür gelten Zahlen, in denen Dezimalstellen durch einen Punkt wiedergegeben und Kommata als Tausendertrennzeichen verwendet werden.

Punkt. Als gültige Werte hierfür gelten Zahlen, in denen Dezimalstellen durch ein Komma wiedergegeben und Punkte als Tausendertrennzeichen verwendet werden.

Anmerkung: Werte, die unzulässige Zeichen für das gewählte Format enthalten, werden als fehlende Werte behandelt. Werte, in denen eines der angegebenen Trennzeichen enthalten ist, werden als Mehrfachwerte behandelt.

Text-Assistent: Schritt 6

Abbildung 3-22
Text-Assistent: Schritt 6



Dies ist der letzte Schritt im Text-Assistenten. Sie können Ihre Einstellungen in einer Datei speichern, um sie beim Importieren ähnlicher Textdatendateien verwenden zu können. Sie können auch die vom Text-Assistenten erzeugte Syntax in ein Syntax-Fenster einfügen. Sie können die Syntax dann anpassen und/oder speichern, um sie bei anderen Sitzungen oder Produktionsjobs einsetzen zu können.

Daten lokal zwischenspeichern. Ein Zwischenspeicher (Cache) für die Daten ist eine vollständige Kopie der Datendatei, die temporär auf der Festplatte gespeichert wird. Zwischenspeichern der Datendatei kann die Leistung verbessern.

Einlesen von Daten aus Dimensions

Sie können Daten aus SPSS Dimensions-Produkten wie Quanvert, Quancept und mrInterview einlesen.

Wenn Sie auf Dimensions-Datenquellen zugreifen möchten, müssen folgende Komponenten auf dem Computer installiert sein, auf dem SPSS ausgeführt wird:

- .NET Framework
- Dimensions-Datenmodell und OLE DB Access

Die Versionen dieser Komponenten, die mit dieser Version von SPSS kompatibel sind, können von der Installations-CD von SPSS installiert werden. Sie sind über das automatisch angezeigte Menü verfügbar. Dimensions-Datenquellen können nur im Modus für lokale Analysen eingelesen werden. Diese Funktion ist im Modus für verteilte Analyse mit dem SPSS-Server nicht verfügbar.

So lesen Sie Daten aus einer Dimensions-Datenquelle ein:

▶ Wählen Sie in einem beliebigen SPSS-Fenster die folgenden Befehle aus den Menüs aus:

Datei
  Dimensionsdaten öffnen

▶ Geben Sie im Dialogfeld "Eigenschaften der Datenverknüpfung" auf der Registerkarte "Verbindung" die Metadatendatei, den Falldatentyp und die Falldatendatei an.

▶ Klicken Sie auf OK.

▶ Wählen Sie im Dialogfeld "Daten aus Dimensions importieren" die gewünschten Variablen und ggf. Fallauswahlkriterien aus.

▶ Klicken Sie auf OK, um die Daten einzulesen.

Registerkarte "Verbindung" im Dialogfeld "Eigenschaften der Datenverknüpfung"

Wenn Sie Daten aus einer Dimensions-Datenquelle einlesen möchten, müssen Sie Folgendes angeben:

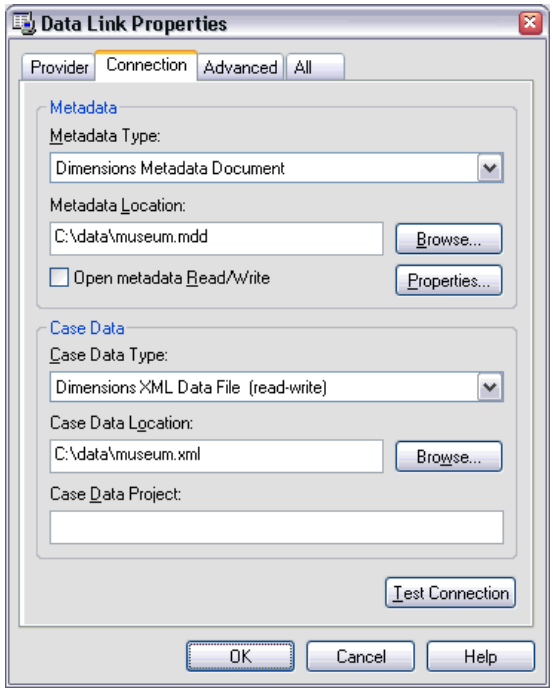
Speicherort der Metadaten. Dies ist die Metadaten-Dokumentdatei (*.mdd*), die Informationen zur Definition der Umfrage enthält.

Falldatentyp. Hierbei handelt es sich um das Format der Falldatendatei. Folgende Formate sind verfügbar:

- Quancept-Datendatei (DRS). Falldaten in einer Quancept-Datei im Format *.drs*, *.drz* oder *.dru*.
- Quanvert-Datenbank. Falldaten in einer Quanvert-Datenbank.
- Dimensions-Datenbank (MS SQL Server). Falldaten in einer relationalen SPSS MR-Datenbank auf einem SQL Server. Diese Option kann verwendet werden, um Daten einzulesen, die mit mrInterview erfasst wurden.
- Dimensions XML-Datendatei. Falldaten in einer XML-Datei.

Speicherort der Falldaten. Dies ist die Datei, die die Falldaten enthält. Das Format dieser Datei muss dem ausgewählten Falldatentyp entsprechen.

Abbildung 3-23
Eigenschaften der Datenverknüpfung: Registerkarte "Verbindung"



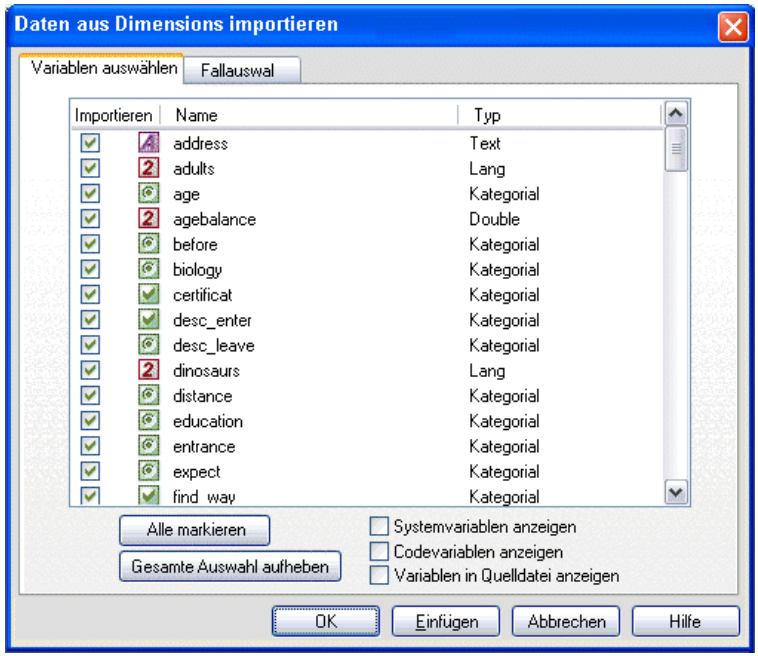
Anmerkung: Das Ausmaß, in welchem die weiteren Einstellungen auf der Registerkarte "Verbindung" sowie auf den weiteren Registerkarten im Dialogfeld "Eigenschaften der Datenverknüpfung" das Einlesen von Daten aus Dimensions in SPSS beeinträchtigen, ist nicht bekannt. Deshalb empfehlen wir, dass Sie keine dieser Einstellungen ändern.

Registerkarte "Variablen auswählen"

Sie können eine Untergruppe von einzulesenden Variablen auswählen. In der Standardeinstellung werden alle Standardvariablen in der Datenquelle angezeigt und ausgewählt.

- **Systemvariablen anzeigen.** Hiermit werden beliebige "Systemvariablen" angezeigt, einschließlich der Variablen, die den Interviewstatus kennzeichnen (*in progress*, *completed*. *finish date* usw.). Sie können dann die gewünschten Systemvariablen auswählen. In der Standardeinstellung werden alle Systemvariablen ausgeschlossen.
- **Codevariablen anzeigen.** Hiermit werden alle Variablen angezeigt, die die Codes darstellen, die für "weitere" Antworten mit offenen Ende für kategoriale Variablen verwendet werden. Sie können dann die gewünschten Codevariablen auswählen. In der Standardeinstellung werden alle Codevariablen ausgeschlossen.
- **Variablen in Quelldatei anzeigen.** Hiermit werden alle Variablen angezeigt, die die Dateinamen der Bilder von eingescannten Antworten enthalten. Sie können dann die gewünschten Variablen der Quelldatei auswählen. In der Standardeinstellung werden alle Variablen der Quelldatei ausgeschlossen.

Abbildung 3-24
Daten aus Dimensions importieren: Registerkarte "Variablen auswählen"



Registerkarte "Fallauswahl"

Bei Dimensions-Datenquellen, die Systemvariablen enthalten, können Sie Fälle auf der Basis einer Anzahl von Kriterien für Systemvariablen auswählen Sie müssen die entsprechenden Systemvariablen nicht in die Liste der einzulesenden Variablen aufnehmen. Die erforderlichen Systemvariablen müssen jedoch in den Quelldaten vorliegen, damit die Auswahlkriterien angewendet werden können. Wenn die erforderlichen Systemvariablen nicht in der Quelldatei vorhanden sind, werden die entsprechenden Auswahlkriterien ignoriert.

Status der Datenerfassung. Sie können Antwortdaten, Testdaten oder beide auswählen. Sie können Fälle auch auf der Basis einer beliebigen Kombination der folgenden Statusparameter für Interviews auswählen:

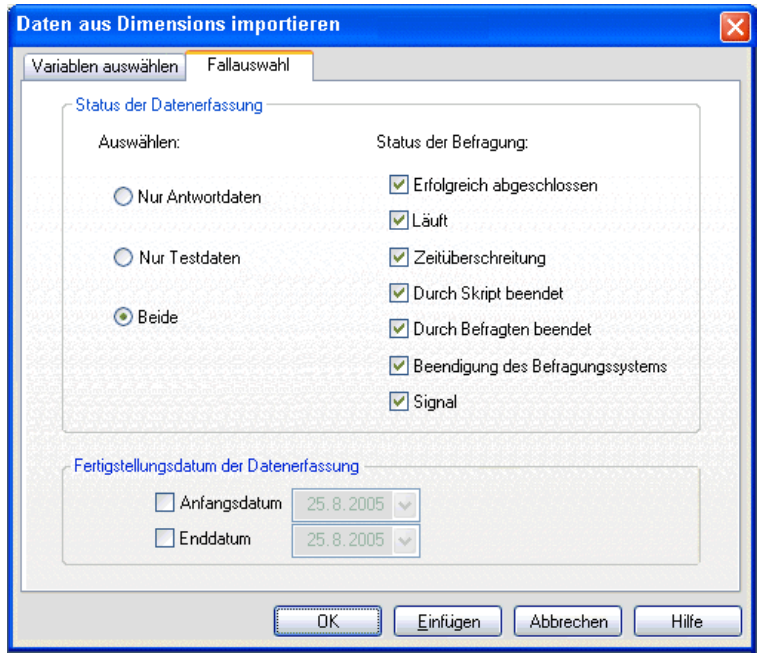
- Erfolgreich abgeschlossen
- Läuft

- Zeitüberschreitung
- Durch Skript beendet
- Durch Befragten beendet
- Beendigung des Befragungssystems
- Signal (durch eine Signalanweisung im Skript beendet)

Fertigstellungsdatum der Datenerfassung. Sie können Fälle auf der Basis des Fertigstellungsdatums der Datenerfassung auswählen.

- **Anfangsdatum.** Es werden Fälle berücksichtigt, bei denen die Datenerfassung am oder nach dem angegebenen Datum abgeschlossen wurde.
- **Enddatum.** Es werden Fälle berücksichtigt, bei denen die Datenerfassung vor dem angegebenen Datum abgeschlossen wurde. Hierbei werden *keine* Fälle berücksichtigt, für die die Datenerfassung am Enddatum abgeschlossen wurde.
- Wenn Sie das Anfangs- als auch das Enddatum angeben, ergibt dies einen Datumsbereich vom Anfangsdatum bis zum Tag vor dem Enddatum.

Abbildung 3-25
Daten aus Dimensions importieren: Registerkarte "Fallauswahl"



Informationen zur Datei

Eine SPSS-Datendatei enthält mehr als nur Rohdaten. Sie enthält außerdem Informationen zu Definitionen von Variablen. Dies umfaßt die folgenden Informationen:

- Variablennamen
- Variablenformate
- Beschreibende Variablen- und Wertelabels

Diese Informationen werden im Datenlexikon einer SPSS-Datendatei gespeichert. Mit dem Daten-Editor können die Informationen zu Definitionen der Variablen eingesehen werden. Es ist außerdem möglich, alle Informationen aus dem Datenlexikon der Arbeitsdatei oder einer beliebigen Datendatei anzuzeigen.

So rufen Sie Informationen über eine Datendatei ab:

- Wählen Sie die folgenden Befehle aus den Menüs im Fenster "Daten-Editor" aus:
  Datei
  Datendatei-Informationen anzeigen
- Wählen Sie für die derzeit geöffnete Datei die Option Arbeitsdatei.
- Wählen Sie für andere Datendateien die Option Externe Datei, und wählen Sie dann die Datendatei aus.

  Die Informationen zur Datendatei werden im Viewer angezeigt.

Speichern von Datendateien

Änderungen, die Sie an einer Datendatei vornehmen, gelten nur für die Dauer einer Sitzung in SPSS. Wenn Sie die Änderungen beibehalten möchten, müssen Sie diese ausdrücklich speichern.

So speichern Sie geänderte Datendateien:

- Aktivieren Sie den Daten-Editor. Klicken Sie dazu auf eine beliebige Stelle des entsprechenden Fensters.
- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Datei
  Speichern

  Die geänderten Daten werden gespeichert. Dabei wird die vorherige Version der Datei überschrieben.

Speichern von Datendateien im Excel-Format

Daten können in drei verschiedenen Microsoft Excel-Dateiformaten gespeichert werden. Die Auswahl des Formats hängt von der Excel-Version ab, die zum Öffnen der Daten verwendet werden soll. Von einer Excel-Anwendung können keine Excel-Dateien geöffnet werden, die in einer neueren Version der Anwendung gespeichert wurden. In Excel 5.0 kann beispielsweise kein Excel 2000-Dokument geöffnet werden. Von Excel 2000 kann jedoch problemlos ein Excel 5.0-Dokument eingelesen werden.

Im Vergleich zum SPSS-Format gibt es für das Excel-Dateiformat einige Einschränkungen. Dazu gehören:

- Variablenbeschreibungen wie fehlende Werte und Variablenlabels werden nicht in die exportierten Excel-Dateien aufgenommen.
- Beim Exportieren in Excel 97 und nachfolgende Versionen gibt es eine Option zum Einschließen von Wertelabels anstelle von Werten.
- Da alle Excel-Dateien auf 256 Datenspalten beschränkt sind, werden nur die ersten 256 Variablen in die exportierte Datei übernommen.
- Dateien für Excel 4.0 und Excel 5.0/95 sind auf 16.384 Datensätze oder Datenzeilen beschränkt. Dateien für Excel 97-2000 können maximal 65.536 Datensätze enthalten. Wenn die Datenmenge diese Beschränkungen überschreitet, wird eine Warnung angezeigt, und die Daten werden bei der für Excel zulässigen Maximalgröße abgeschnitten.

Variablentypen

In der folgenden Tabelle werden die Variablentypen der SPSS-Originaldaten und deren Entsprechungen in den exportierten Excel-Daten dargestellt.

SPSS-Variablentyp	**Excel-Datenformat.**
Numerisch	0.00; #,##0.00; ...
Komma	0.00; #,##0.00; ...
Dollar	$#,##0_); ...
Datum.	t-mmm-jjjj
Zeit	hh:mm:ss
String	Allgemein

Speichern von Datendateien im SAS-Format

Beim Speichern von Daten als SAS-Datei werden verschiedene Aspekte der Daten besonders behandelt. Dazu gehören:

- Bestimmte Zeichen, die für Variablennamen in SPSS zulässig sind, sind in SAS nicht gültig, beispielsweise @, # und $. Diese ungültigen Zeichen werden beim Exportieren der Daten durch einen Unterstrich ersetzt.

- SPSS-Variablennamen, die Mehrbyte-Zeichen enthalten (z. B. japanische oder chinesische Zeichen) werden in Variablenamen der allgemeinen Form *Vnnn* konvertiert, wobei *nnn* ein ganzzahliger Wert ist.
- SPSS-Variablenlabels mit mehr als 40 Zeichen werden beim Exportieren in eine SAS v6-Datei abgeschnitten.
- Sofern SPSS-Variablenlabels vorhanden sind, werden sie den entsprechenden SAS-Variablenlabels zugeordnet. Wenn die SPSS-Daten keine Variablenlabels enthalten, wird dem SAS-Variablenlabel der Variablenname zugeordnet.
- Während in SPSS zahlreiche systemdefinierte fehlende Werte zulässig sind, kann es in SAS nur einen systemdefinierten fehlenden Wert geben. Daher werden alle systemdefinierten fehlenden Werte in den SPSS-Daten nur einem systemdefinierten fehlenden Wert in der SAS-Datei zugeordnet.

Speichern von Wertelabels

Sie haben die Möglichkeit, die der Datendatei zugeordneten Werte und Wertelabels in einer SAS-Syntaxdatei zu speichern. Wenn beispielsweise die Wertelabels für die Datendatei *cars.sav* exportiert werden, enthält die erzeugte Syntaxdatei folgende Zeilen:

```
libname library 'd:\spss\' ;

proc format library = library ;
   value ORIGIN /* Herstellungsland */
      1 = 'Amerika'
      2 = 'Europa'
      3 = 'Japan' ;
   value CYLINDER /* Anzahl der Zylinder */
      3 = '3 Zylinder'
      4 = '4 Zylinder'
      5 = '5 Zylinder'
      6 = '6 Zylinder'
      8 = '8 Zylinder'
   value FILTER__ /* zylinder = 1 | zylinder = 2 (FILTER) */
      0 = 'Nicht ausgewählt'
```

```
      1 = 'Ausgewählt'

proc datasets library = library ;
modify cars;
   format    ORIGIN ORIGIN.;
   format  CYLINDER CYLINDER.;
   format  FILTER__ FILTER__.;
quit;
```

Diese Funktion wird für SAS-Transportdateien nicht unterstützt.

Variablentypen

In der folgenden Tabelle werden die Variablentypen der SPSS-Originaldaten und deren Entsprechungen in den exportierten SAS-Daten dargestellt.

SPSS-Variablentyp	SAS-Variablentyp	SAS-Datenformat
Numerisch	Numerisch	12
Komma	Numerisch	12
Punkt	Numerisch	12
Wissenschaftliche Notation	Numerisch	12
Datum.	Numerisch	(Datum), z. B. MMDDYY10, ...
Datum (Uhrzeit)	Numerisch	Time18
Dollar	Numerisch	12
Spezielle Währung	Numerisch	12
String	Zeichen	$8

Speichern von Datendateien im Stata-Format

- Die Daten können im Format von Stata Versionen 5-8 geschrieben werden, sowohl im Intercooled- als auch im SE-Format (nur Versionen 7-8).
- Die Datendateien, die im Format von Stata 5 gespeichert wurden, können von Stata 4 eingelesen werden.

- Die ersten 80 Byte von Variablenlabels werden als Stata-Variablenlabels gespeichert.
- Bei numerischen Variablen werden die ersten 80 Byte der Variablenlabels als Stata-Wertelabels gespeichert. Bei String-Variablen werden die Wertelabels verworfen.
- Beim Format der Versionen 7 und 8 werden die ersten 32 Byte der Variablennamen unter Berücksichtigung der Groß- und Kleinschreibung als Stata-Variablennamen gespeichert. Beim Format von früheren Versionen werden die ersten 8 Byte der Variablennamen als Stata-Variablennamen gespeichert. Alle Zeichen außer Buchstaben, Ziffern und Unterstrichen werden in Unterstriche umgewandelt.
- SPSS-Variablennamen, die Mehrbyte-Zeichen enthalten (z. B. japanische oder chinesische Zeichen) werden in Variablenamen der allgemeinen Form *Vnnn* konvertiert, wobei *nnn* ein ganzzahliger Wert ist.
- Beim Format der Versionen 5-6 und den Intercooled-Versionen 7-8 werden die ersten 80 Byte der String-Werte gespeichert. Beim Format von Stata SE 7-8 werden die ersten 244 Byte der String-Werte gespeichert.
- Beim Format der Versionen 5-6 und den Intercooled-Versionen 7-8 werden nur die ersten 2.047 Variablen gespeichert. Beim Format von Stata SE 7-8 werden nur die ersten 32.767 Variablen gespeichert.

SPSS-Variablentyp	**Stata-Variablentyp**	**Stata-Datenformat**
Numerisch	Numeric	g
Komma	Numeric	g
Punkt	Numeric	g
Wissenschaftliche Notation	Numeric	g
Datum, Datum/Zeit	Numeric	D_m_Y
Zeit, DTime	Numeric	g (Anzahl der Sekunden)
Wkday	Numeric	g (1-7)
Moyr	Numeric	g (1-12)
Dollar	Numeric	g
Spezielle Währung	Numeric	g
String	String	s

Speichern von Datendateien in anderen Formaten

- ▶ Aktivieren Sie den Daten-Editor. Klicken Sie dazu auf eine beliebige Stelle des entsprechenden Fensters.
- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:
  Datei
  Speichern unter…
- ▶ Wählen Sie aus der Dropdown-Liste einen Dateityp aus.
- ▶ Geben Sie einen Namen für die neue Datendatei ein.

So schreiben Sie die Variablennamen in die erste Zeile eines Tabellenkalkulationsblatts oder die erste Zeile einer Tabulator-getrennten Textdatei:

- ▶ Klicken Sie im Dialogfeld "Daten speichern unter" auf Variablennamen im Arbeitsblatt speichern.

So speichern Sie Wertelabels anstelle von Datenwerten im Format von Excel 97:

- ▶ Klicken Sie im Dialogfeld "Daten speichern unter" auf Sofern definiert, Wertelabels statt Datenwerte speichern.

So speichern Sie Wertelabels in einer SAS-Syntaxdatei (nur nach Auswahl eines SAS-Dateityps aktiv):

- ▶ Klicken Sie im Dialogfeld "Daten speichern unter" auf Wertelabels in einer SAS-Datei speichern.

Speichern von Daten: Datendateitypen

Sie können Ihre Daten in den folgenden Formaten speichern:

SPSS (*.sav). SPSS-Format.

- Programmversionen vor Version 7.5 können keine Datendateien lesen, die im SPSS-Format gespeichert werden.
- Bei der Verwendung von Datendateien mit Variablennamen mit mehr als 8 Byte in SPSS 10.x oder 11.x werden eindeutige 8 Byte umfassende Versionen der Variablen verwendet. Die ursprünglichen Variablennamen bleiben jedoch für die

Verwendung in Version 12.0 oder höher erhalten. Bei Versionen vor SPSS 10.0 gehen die ursprünglichen langen Variablennamen beim Speichern der Datendatei verloren.

- Wenn Sie Datendateien mit String-Variablen mit mehr als 255 Byte in SPSS-Versionen vor Version 13.0 verwenden, werden diese String-Variablen in mehrere String-Variablen mit je 255 Byte aufgeteilt.

SPSS 7.0 (*.sav). Dateien im Format SPSS 7.0 für Windows. Datendateien, die im Format von SPSS 7.0 gespeichert wurden, können von SPSS 7.0 und früheren SPSS-Versionen für Windows eingelesen werden. Sie enthalten jedoch keine Definitionen von Mehrfachantworten-Sets oder Informationen aus Data Entry für Windows.

SPSS/PC+ (*.sys). Dateien im Format von SPSS/PC+. Wenn die Datendatei mehr als 500 Variablen enthält, werden nur die ersten 500 gespeichert. Bei Variablen mit mehr als einem benutzerdefinierten fehlenden Wert werden die zusätzlichen benutzerdefinierten fehlenden Werte in den ersten benutzerdefinierten fehlenden Wert umkodiert.

SPSS portable (*.por). Portable SPSS-Dateien, die von anderen Versionen von SPSS und Versionen unter anderen Betriebssystemen eingelesen werden können, also beispielsweise auf dem Macintosh oder unter UNIX. Variablennamen sind auf 8 Byte begrenzt und werden gegebenenfalls automatisch in eindeutige 8 Byte umfassende Namen konvertiert.

Tabulator-getrennt (*.dat). ASCII-Textdateien, bei denen die Werte durch Tabulatoren getrennt sind.

Festes ASCII (*.dat). ASCII-Textdateien im festen Format. Hierbei werden für alle Variablen die Standard-Schreibformate verwendet. Zwischen den Feldern der Variablen befinden sich weder Tabulator- noch Leerzeichen.

Excel 2.1 (*.xls). Tabellenkalkulationsdateien im Format von Microsoft Excel 2,1. Die Dateien dürfen höchstens 256 Variablen und 16.384 Zeilen enthalten.

Excel 97 und nachfolgende Versionen (*.xls). Tabellenkalkulationsdateien im Format von Microsoft Excel 97/2000/XP. Die Dateien dürfen höchstens 256 Variablen und 65.536 Zeilen enthalten.

1-2-3 Version 3.0 (*.wk3). Tabellenkalkulationsdateien im Format von Lotus 1-2-3 Version 3,0. Es können höchstens 256 Variablen gespeichert werden.

1-2-3 Version 2.0 (*.wk1). Tabellenkalkulationsdateien im Format von Lotus 1-2-3 Version 2.0. Es können höchstens 256 Variablen gespeichert werden.

1-2-3 Version 1.0 (*.wks). Tabellenkalkulationsdateien im Format von Lotus 1-2-3 Version 1A. Es können höchstens 256 Variablen gespeichert werden.

SYLK (*.slk). Dateien im "Symbolic Link"-Format für Tabellenkalkulationsdateien von Microsoft Excel und Multiplan. Es können höchstens 256 Variablen gespeichert werden.

dBASE IV (*.dbf). dBASE IV-Format.

dBASE III (*.dbf). dBASE III-Format.

dBASE II (*.dbf). dBASE II-Format.

SAS v6 für Windows (*.sd2). Dateien im Format SAS V6 für Windows/OS2.

SAS v6 für UNIX (*.ssd01). Dateien im Format SAS V6 für UNIX (Sun, HP, IBM).

SAS v6 für Alpha/OSF (*.ssd04). Dateiformat SAS V 6 für Alpha/OSF (DEC UNIX).

SAS v7+ für Windows, kurze Erweiterung (*.sd7). SAS Versionen 7-8 für Windows, kurzes Dateinamensformat.

SAS v7+ für Windows, lange Erweiterung (*.sas7bdat). SAS Versionen 7-8 für Windows, langes Dateinamensformat.

SAS v7+ für UNIX (*.ssd01). SAS v8 für UNIX.

SAS Transport (*.xpt). SAS-Transportdatei.

Stata Versionen 4-5 (*.dta).

Stata Version 6 (*.dta).

Stata Version 7 Intercooled (*.dta).

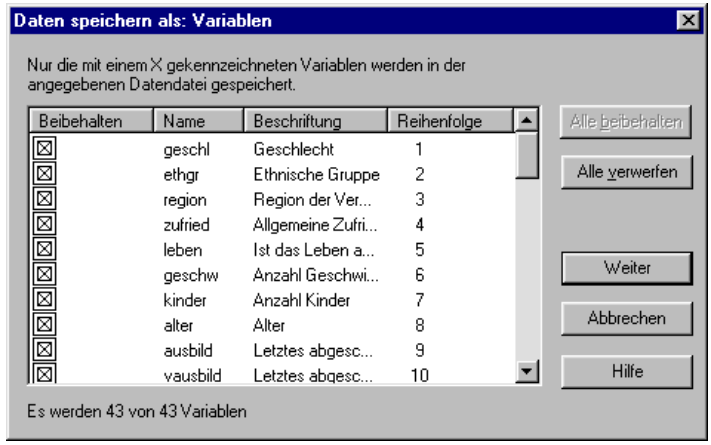
Stata Version 7 SE (*.dta).

Stata Version 8 Intercooled (*.dta).

Stata Version 8 SE (*.dta).

Speichern von Untergruppen von Variablen

Abbildung 3-26
Dialogfeld "Daten speichern als: Variablen"



Im Dialogfeld "Daten speichern als: Variablen" können Sie die Variablen auswählen, die in der neuen Datendatei gespeichert werden sollen. In der Standardeinstellung werden alle Variablen gespeichert. Heben Sie die Auswahl für die Variablen auf, die nicht gespeichert werden sollen, oder klicken Sie auf Alle verwerfen, und wählen Sie dann die zu speichernden Variablen aus.

So speichern Sie eine Untergruppe von Variablen

- Aktivieren Sie den Daten-Editor. Klicken Sie dazu auf eine beliebige Stelle des entsprechenden Fensters.
- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Datei
  Speichern unter…
- Wählen Sie den der Datei entsprechenden Dateityp aus.
- Klicken Sie auf Variablen.
- Wählen Sie die Variablen aus, die Sie speichern möchten.

Datei speichern: Optionen

Bei Tabellenkalkulationsdateien und Textdateien mit Tabulatoren als Trennzeichen können die Variablennamen in die erste Zeile der Datei geschrieben werden.

Schützen der ursprünglichen Daten

Um eine versehentliche Änderung/Löschung der ursprünglichen Daten zu verhindern, können Sie die Datei mit einem Schreibschutz versehen.

- Wählen Sie die folgenden Menübefehle des Daten-Editors aus:

  Datei
  Datei als schreibgeschützt markieren

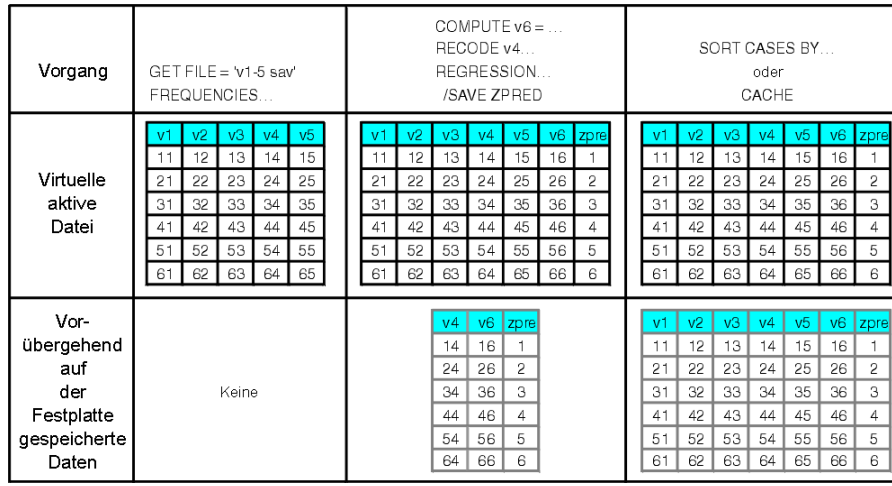
Wenn Sie danach Änderungen an den Daten vornehmen und versuchen, die Datendatei zu speichern, können Sie die Daten nur unter einem anderen Dateinamen speichern, sodass die ursprünglichen Daten unverändert erhalten bleiben.

Sie können die Dateiberechtigungen wieder auf "Lesen/Schreiben" zurücksetzen, indem Sie im Menü "Datei" die Option Datei für Lese-/Schreibzugriff markieren auswählen.

Virtuelle aktive Datei

Eine virtuelle aktive Datei ermöglicht die Arbeit mit großen Datendateien, ohne daß dabei mindestens so viel temporärer Speicherplatz auf der Festplatte vorhanden sein muß, wie die Größe der Datendatei beträgt. Bei den meisten Analyse- und Diagrammprozeduren wird die ursprüngliche Datenquelle bei jedem Ausführen erneut eingelesen. Bei Prozeduren, bei denen die Daten modifiziert werden, wird eine gewisse Menge an temporärem Speicherplatz auf der Festplatte benötigt, um die Änderungen aufzuzeichnen, und bei einigen Vorgängen ist immer mindestens ausreichend Speicherplatz auf der Festplatte für eine vollständige Kopie der Datendatei erforderlich.

Abbildung 3-27
Anforderungen an temporären Speicherplatz auf der Festplatte



Vorgänge, bei denen kein temporärer Speicherplatz auf der Festplatte benötigt wird:

- Einlesen von SPSS-Datendateien
- Zusammenfügen von zwei oder mehr SPSS-Datendateien
- Einlesen von Datenbankdateien mit dem Datenbank-Assistenten
- Zusammenfügen einer SPSS-Datendatei mit einer Datenbanktabelle
- Ausführen von Prozeduren, die Daten einlesen (beispielsweise Häufigkeiten, Kreuztabellen und Explorative Datenanalyse)

Vorgänge, bei denen mindestens eine Datenspalte in temporärem Speicherplatz auf der Festplatte erstellt wird:

- Berechnen von neuen Variablen
- Umkodieren von vorhandenen Variablen
- Ausführen von Prozeduren, bei denen Variablen erstellt oder modifiziert werden (beispielsweise das Speichern von vorhergesagten Werten bei der linearen Regression)

Vorgänge, bei denen eine vollständige Kopie der Datendatei in temporärem Speicherplatz auf der Festplatte erstellt wird:

- Einlesen von Excel-Dateien
- Ausführen von Prozeduren zum Sortieren von Daten (beispielsweise die Prozeduren "Fälle sortieren" und "Datei aufteilen")
- Einlesen von Daten mit den Syntaxbefehlen `GET TRANSLATE` und `DATA LIST`
- Verwenden der Funktionen für das Ablegen von Daten im Zwischenspeicher oder des Syntaxbefehls `CACHE`
- Starten von anderen Anwendungen aus SPSS heraus, welche die Datendatei einlesen (beispielsweise AnswerTree und DecisionTime)

Anmerkung: Der Syntaxbefehl `GET DATA` stellt ähnliche Funktionen wie der Syntaxbefehl `DATA LIST` bereit, erstellt jedoch keine vollständige Kopie der Datendatei in temporärem Speicherplatz auf der Festplatte. Mit dem Syntaxbefehl `SPLIT FILE` werden die Daten in der Datendatei nicht sortiert. Deshalb wird auch keine Kopie der Datendatei erstellt. Damit dieser Befehl jedoch ordnungsgemäß ausgeführt werden kann, müssen die Daten sortiert sein. Über die Benutzeroberfläche des Dialogfelds dieser Prozedur wird die Datendatei automatisch sortiert und eine vollständige Kopie der Datendatei erstellt. (Die Befehlssyntax ist in der Studentenversion nicht verfügbar.)

Vorgänge, bei denen in der Standardeinstellung eine vollständige Kopie der Datendatei erstellt wird:

- Einlesen von Datenbanken mit dem Datenbank-Assistenten
- Einlesen von Textdateien mit dem Text-Assistenten

Der Text-Assistent bietet eine optionale Einstellung zum automatischen Zwischenspeichern der Daten. Diese Option ist standardmäßig ausgewählt. Sie können diese Auswahl aufheben, indem Sie das Kontrollkästchen Daten in lokalen Zwischenspeicher deaktivieren. Beim Databank-Assistenten können Sie die erstellte Befehlssyntax einfügen und den Befehl `CACHE` löschen.

Erstellen eines Zwischenspeichers für Daten

Die virtuelle aktive Datei kann die benötigte Menge an temporärem Speicherplatz auf der Festplatte drastisch reduzieren. Das Nichtvorhandensein einer temporären Kopie der eigentlich aktiven Datei bedeutet aber auch, daß die ursprüngliche Datendatei für jede Prozedur neu eingelesen werden muß. Bei großen Datendateien, die aus einer externen Quelle eingelesen werden, kann das Erstellen einer temporären Kopie der Daten die Leistung steigern. Bei Tabellen in einer Datenbank beispielsweise muß die SQL-Abfrage, mit der die Informationen aus der Datenbank ausgelesen werden, für jeden Befehl und jede Prozedur erneut ausgeführt werden, bei denen Daten eingelesen werden. Da fast alle Statistik- und Diagrammprozeduren die Daten einlesen müssen, wird die SQL-Abfrage für jede aufgerufene Prozedur erneut ausgeführt. Bei einer großen Anzahl an Prozeduren kann dies zu einer beträchtlichen Steigerung der für die Verarbeitung benötigten Zeit führen.

Wenn auf dem Computer, auf dem die Analyse durchgeführt wird (der lokale Computer oder der Remote-Server), ausreichend Speicherplatz auf der Festplatte vorhanden ist, können Sie die mehrfache Ausführung von SQL-Afragen vermeiden und somit die Verarbeitungszeit verringern, indem Sie einen Zwischenspeicher für die Daten aus der aktiven Datei anlegen. Der Zwischenspeicher für die Daten ist eine temporäre Kopie der gesamten Daten.

Anmerkung: In der Standardeinstellung erstellt der Datenbank-Assistent automatisch einen Zwischenspeicher für die Daten. Wenn Sie aber mit Hilfe des Syntaxbefehls `GET DATA` eine Datenbank einlesen, wird nicht automatisch ein Zwischenspeicher für die Daten erstellt. (Die Befehlssyntax ist in der Studentenversion nicht verfügbar.)

So erstellen Sie einen Zwischenspeicher für die Daten

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Datei
  Daten in Zwischenspeicher...

- Klicken Sie auf OK oder Jetzt zwischenspeichern.

Mit OK werden die Daten in den Zwischenspeicher überführt, wenn das Programm die Daten das nächste Mal einliest, beispielsweise beim nächsten Ausführen einer statistischen Prozedur. In der Regel empfiehlt sich dieses Vorgehen, da hierbei kein zusätzlicher Aufwand beim Einlesen der Daten entsteht. Mit Jetzt zwischenspeichern werden die Daten sofort in den Zwischenspeicher übertragen. In den meisten

Situationen ist dies nicht notwendig. Jetzt zwischenspeichern ist primär aus zwei Gründen nützlich:

- Eine Datenquelle ist gesperrt und kann nicht durch andere aktualisiert werden, bis Sie Ihre Sitzung beenden, eine andere Datenquelle öffnen oder die Daten zwischenspeichern.
- Bei umfangreichen Datenquellen erfolgt der Bildlauf durch den Inhalt der Registerkarte "Datenansicht" des Daten-Editors viel schneller, wenn Sie die Daten zwischenspeichern.

So können Sie Daten automatisch zwischenspeichern

Mit Hilfe des Befehls `SET` können Sie nach einer festgelegten Anzahl von Änderungen in der aktiven Datendatei automatisch einen Zwischenspeicher für die Daten erstellen. In der Standardeinstellung wird die aktive Datendatei nach 20 Änderungen automatisch zwischengespeichert.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Datei
  Neu
    Syntax

▶ Geben Sie im Syntax-Fenster folgendes ein: SET CACHE n. (Dabei steht *n* für die Anzahl der Änderungen in der aktiven Datendatei, nach der die Datendatei zwischengespeichert wird.)

▶ Wählen Sie die folgenden Befehle aus den Menüs des Syntax-Fensters aus:

Ausführen
  Alles

Anmerkung: Die Einstellung für den Zwischenspeicher wird nicht für alle Sitzungen übernommen. Bei jedem Start einer neuen Sitzung wird der Wert auf den Standardwert 20 zurückgesetzt.

Kapitel
4

Modus für verteilte Analysen

Beim Modus für verteilte Analysen können Sie speicherintensive Vorgänge von Ihrem lokalen Computer weg auf einen anderen Computer auslagern. Da die für verteilte Analysen eingesetzten Remote-Server in der Regel leistungsfähiger und schneller als Ihr lokaler Computer sind, kann ein sinnvoller Einsatz des Modus für verteilte Analysen die für die Verarbeitung benötigte Zeit beträchtlich verringern. In den folgenden Situationen kann die verteilte Analyse auf einem Remote-Server nützlich sein:

- Sie arbeiten mit umfangreichen Datendateien oder mit Daten, die aus einer Datenbank eingelesen werden.
- Für die Analyse müssen speicherintensive Aufgaben durchgeführt werden. Alle Aufgaben, deren Verarbeitung im Modus für lokale Analysen sehr lange dauert, können möglicherweise von der verteilten Analyse profitieren.

Die verteilte Analyse greift nur bei datenbezogenen Aufgaben. Hierzu gehören beispielsweise das Einlesen von Daten, das Transformieren von Daten, das Berechnen neuer Variablen und das Berechnen von Statistiken. Eine verteilte Analyse wirkt sich nicht auf Aufgaben in Zusammenhang mit der Bearbeitung von Ausgaben aus. Hierzu gehören beispielsweise das Bearbeiten von Pivot-Tabellen oder das Modifizieren von Diagrammen.

Anmerkung: Der Modus für verteilte Analysen steht nur zur Verfügung, wenn sowohl eine lokale Version von SPSS als auch der Zugriff auf eine lizenzierte Server-Version von SPSS vorliegt, die auf einem Remote-Server installiert ist.

Verteilte Analyse im Vergleich zur lokalen Analyse

Im Folgenden finden Sie einige Richtlinien für die Wahl des Modus für verteilte Analysen und für lokale Analysen:

Datenbankzugriff. Jobs, bei denen Datenbankabfragen ausgeführt werden, laufen möglicherweise schneller im verteilten Modus, wenn der Server über privilegierten Zugriff auf die Datenbank verfügt oder auf demselben Computer wie die Datenbank-Software ausgeführt wird. Wenn die erforderliche Software für den Datenbankzugriff darüber hinaus nur auf dem Server verfügbar ist oder Ihr Netzwerkadministrator Ihnen das Herunterladen von umfangreichen Datentabellen untersagt, können Sie nur im verteilten Modus auf die Datenbank zugreifen.

Verhältnis von Berechnungen zu Ausgabe. Am meisten von der verteilten Analyse profitieren Befehle, bei denen umfangreiche Berechnungen durchgeführt, aber nur wenige Ausgaben produziert werden, beispielsweise wenige und kleine Pivot-Tabellen, kurze Textausgaben oder wenige und einfache Diagramme. Die Durchsatzsteigerung hängt hauptsächlich von der Rechenleistung des Remote-Servers ab.

Kleine Jobs. Jobs, die im lokalen Modus schnell ausgeführt werden, laufen im verteilten Modus aufgrund der bei der Client-/Server-Architektur anfallenden Mehrverabeitung fast immer langsamer.

Diagramme. Für fallorientierte Diagramme, beispielsweise Streudiagramme, Residuen-Diagramme bei der Regression und Sequenzdiagramme, sind Rohdaten auf dem lokalen Computer erforderlich. Bei umfangreichen Datendateien oder Datenbanktabellen kann dieser Prozess den Durchsatz im verteilten Modus verringern, weil die Daten vom Remote-Server an den lokalen Computer übermittelt werden müssen. Bei anderen Diagrammen, die auf zusammengefaßten oder aggregierten Daten basieren, fällt der Durchsatz zufriedenstellender aus, weil die Daten auf dem Server aggregiert werden.

Interaktive Grafiken. Wenn die Rohdaten mit einer interaktiven Grafik gespeichert werden (dies ist eine optionale Einstellung), müssen möglicherweise große Datenmengen vom Remote-Server an den lokalen Computer übertragen werden. Dies kann die zum Speichern der Ergebnisse benötigte Zeit beträchtlich steigern.

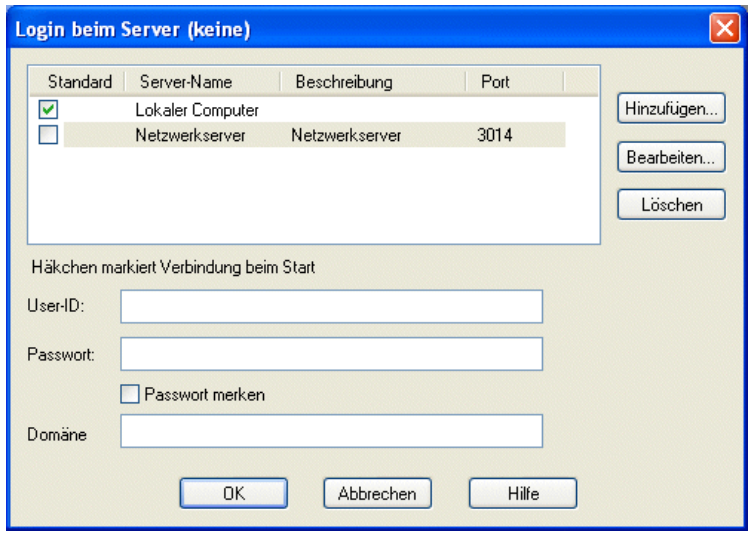
Pivot-Tabellen. Im verteilten Modus dauert das Erstellen von umfangreichen Pivot-Tabellen möglicherweise länger. Dies gilt insbesondere für die Prozedur "OLAP-Würfel" und Tabellen mit den Daten einzelner Fälle, beispielsweise die Tabellen in der Prozedur "Zusammenfassen".

Textausgabe. Je mehr Text produziert wird, desto langsamer erfolgt die Ausgabe im verteilten Modus, weil der Text auf dem Remote-Server produziert und anschließend für die Anzeige auf den lokalen Computer übertragen wird. Bei der Textausgabe fällt jedoch nur wenig Mehrverarbeitung an, so daß Text in der Regel schnell übertragen wird.

Login beim SPSS-Server

Im Dialogfeld "Login beim Server" können Sie auswählen, welcher Computer Befehle verarbeiten und Prozeduren ausfühhren soll. Sie können Ihren lokalen Computer oder einen Remote-Server auswählen.

Abbildung 4-1
Dialogfeld "Login beim SPSS-Server"



Sie können der Liste Remote-Server hinzufügen, Remote-Server in der Liste ändern und aus dieser entfernen. Für Remote-Server ist in der Regel ein Benutzername und ein Passwort erforderlich. Möglicherweise müssen Sie außerdem einen

Domänennamen angeben. Wenden Sie sich an Ihren Systemadministrator, um Informationen zu verfügbaren Servern, Benutzernamen, Paßwörtern, Domänennamen und andere Verbindungsinformationen zu erhalten.

Sie können einen Standard-Server auswählen und die Benutzernamen, Domänennamen und Passwörter für beliebige Server speichern. Beim Starten einer neuen Sitzung wird automatisch eine Verbindung mit dem Standard-Server hergestellt.

Hinzufügen und Bearbeiten von Einstellungen für die Server-Anmeldung

Verwenden Sie das Dialogfeld “Einstellungen für Server-Anmeldung”, um Verbindungsinformationen für Remote-Server im Modus für verteilte Analysen hinzuzufügen bzw. zu bearbeiten.

Abbildung 4-2
Dialogfeld “Einstellungen für Server-Anmeldung”

Wenden Sie sich an Ihren Systemadministrator, um Informationen zu den verfügbaren Servern, Portnummern für diese Server und weitere Verbindungsinformationen zu erhalten. Verwenden Sie Secure Socket Layer nur, wenn Sie von Ihrem Administrator dazu angewiesen wurden.

Server-Name. Der Name des Servers kann ein dem Computer zugewiesener alphanumerischer Name (beispielsweise “NetzwerkServer”) oder eine dem Computer zugewiesene eindeutige IP-Adresse sein (beispielsweise 202.123.456.78).

Portnummer. Die Portnummer bezeichnet den Port, den die Serversoftware für die Kommunikation verwendet.

Beschreibung. Sie können eine optionale Beschreibung eingeben, die in der Serverliste angezeigt werden soll.

Stellen Sie eine Verbindung mit Secure Socket Layer her. Secure Socket Layer (SSL) verschlüsselt Anforderungen für verteilte Analysen, wenn diese an den SPSS-Remote-Server gesendet werden. Verwenden Sie SSL nicht, ohne zuvor mit Ihrem Administrator Rücksprache gehalten zu haben. SSL muss auf Ihrem Desktop-Computer und auf dem Server konfiguriert sein, damit diese Option aktiviert werden kann.

So wählen Sie einen Server aus, wechseln den Server oder fügen einen neuen Server hinzu:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Datei
  Server umschalten...

So wählen Sie einen Standardserver aus:

- Wählen Sie in der Serverliste das Kästchen neben dem Server aus, den Sie verwenden möchten.
- Geben Sie den Benutzernamen, Domänennamen und das Passwort ein, die Sie vom Administrator erhalten haben.

  Anmerkung: Beim Starten einer neuen Sitzung wird automatisch eine Verbindung mit dem Standard-Server hergestellt.

So schalten Sie auf einen anderen Server um:

- Wählen Sie einen Server aus der Liste aus.
- Geben Sie erforderlichenfalls Ihren Benutzernamen, den Domänennamen und Ihr Paßwort ein.

  Anmerkung: Wenn Sie in einer Sitzung den Server wechseln, werden alle geöffneten Fenster geschlossen. Ehe die Fenster geschlossen werden, werden Sie aufgefordert, die vorgenommenen Änderungen zu speichern.

So fügen Sie einen Server hinzu:

- Besorgen Sie sich die Verbindungsinformationen für den Server vom Administrator.

- Klicken Sie auf Hinzufügen, um das Dialogfeld “Einstellungen für Server-Anmeldung” zu öffnen.
- Geben Sie die Verbindungsinformationen und optionalen Einstellungen ein, und klicken Sie anschließend auf OK.

So bearbeiten Sie einen Server:

- Wenden Sie sich an den Administrator, um die geänderten Verbindungsinformationen für den Server zu erhalten.
- Klicken Sie auf Bearbeiten, um das Dialogfeld “Einstellungen für Server-Anmeldung” zu öffnen.
- Geben Sie die Änderungen ein, und klicken Sie anschließend auf OK.

Öffnen von Datendateien auf einem Remote-Server

Abbildung 4-3
Dialogfeld “Open Remote File”

Im Modus für verteilte Analysen wird statt des Standard-Dialogfelds “Datei öffnen” das Dialogfeld “Entfernte Datei öffnen” angezeigt.

- Die Inhalte in der Liste der verfügbaren Dateien, Ordner und Laufwerke hängen davon ab, was auf dem Remote-Server bzw. vom Remote-Server aus verfügbar ist. Der Name des aktuellen Servers wird im oberen Teil des Dialogfelds angezeigt.
- Im Modus für verteilte Analysen können Sie nur auf Dateien auf dem lokalen Computer zugreifen, wenn Sie das Laufwerk oder die Ordner mit den Datendateien für den gemeinsamen Zugriff freigeben.
- Wird auf dem Server ein anderes Betriebssystem ausgeführt (Ihr Computer läuft beispielsweise unter Windows und der Server unter UNIX), werden Sie im Modus für verteilte Analysen wahrscheinlich keinen Zugriff auf lokale Datendateien haben, selbst wenn sie sich in freigegebenen Ordnern befinden.

So öffnen Sie Datendateien auf dem Remote-Server:

▶ Wenn noch keine Verbindung mit dem Remote-Server hergestellt wurde, melden Sie sich beim Remote-Server an.

▶ Wählen Sie je nach Typ der gewünschten Datendatei die folgenden Befehle aus den Menüs aus:

Datei
  Öffnen
    Daten...

oder

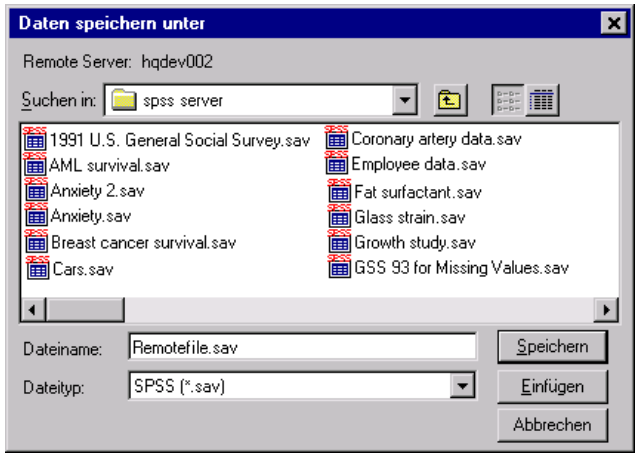
Datei
  Datenbank öffnen

oder

Datei
  Textdaten einlesen...

Speichern von Datendateien auf dem Remote-Server

Abbildung 4-4
Dialogfeld "Entfernte Datei speichern"



Im Modus für verteilte Analysen wird statt des Standard-Dialogfelds "Datei speichern" das Dialogfeld "Entfernte Datei speichern" angezeigt.

Die Inhalte in der Liste der verfügbaren Ordner und Laufwerke hängen davon ab, was auf dem Remote-Server bzw. vom Remote-Server aus verfügbar ist. Der Name des aktuellen Servers wird im oberen Teil des Dialogfelds angezeigt. Sie können nur auf Ordner auf dem lokalen Computer zugreifen, wenn Sie das Laufwerk oder die Ordner für den gemeinsamen Zugriff freigeben. Wird auf dem Server ein anderes Betriebssystem ausgeführt (Ihr Computer läuft beispielsweise unter Windows und der Server unter UNIX), werden Sie im Modus für verteilte Analysen wahrscheinlich keinen Zugriff auf lokale Datendateien haben, selbst wenn sie sich in freigegebenen Ordnern befinden. Wenn Datendateien in einem lokalen Ordner gespeichert werden sollen, müssen diese mit Schreibberechtigung versehen werden.

So speichern Sie Datendateien auf einem Remote-Server:

- Aktivieren Sie das Fenster des Daten-Editors.
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Datei
  Speichern (oder Speichern unter...)

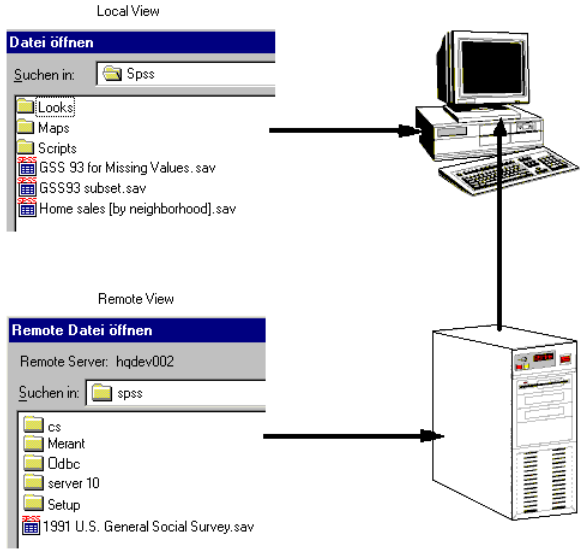
Zugriff auf Datendateien im Modus für lokale und für verteilte Analysen Modalwert

Es hängt von dem Computer ab, den Sie zum Verarbeiten von Befehlen und Ausführen von Prozeduren verwenden, welche Datendateien, Ordner (Verzeichnisse) und Laufwerke sowohl auf dem lokalen Computer als auch im Netzwerk angezeigt werden. Denken Sie daran, dass es sich bei diesem Computer nicht notwendigerweise um den Computer handelt, an dem Sie arbeiten.

Modus für lokale Analysen. Wenn Sie Ihren lokalen Computer als "Server" einsetzen, werden im Dialogfeld zum Öffnen von Dateien dieselben Datendateien, Ordner und Laufwerke angezeigt wie in anderen Anwendungen oder im Windows-Explorer. Es werden alle Datendateien und Ordner auf Ihrem Computer sowie alle Dateien und Ordner auf verbundenen Netzlaufwerken angezeigt.

Modus für verteilte Analysen. Wenn Sie einen anderen Computer als Remote-Server zum Verarbeiten von Befehlen und Ausführen von Prozeduren einsetzen, werden Datendateien, Ordner und Laufwerke aus der Sicht des Remote-Servers angezeigt. Möglicherweise werden bekannte Ordnernamen wie *Programme* und Laufwerke wie *C:* angezeigt. Hierbei handelt es sich aber *nicht* um Ordner und Laufwerke auf dem lokalen Computer, sondern um Ordner und Dateien auf dem Remote-Server.

Abbildung 4-5
Lokale und Remote-Ansichten



Im Modus für verteilte Analysen können Sie nur auf Datendateien auf dem lokalen Computer zugreifen, wenn Sie das Laufwerk oder die Ordner mit den Datendateien für den gemeinsamen Zugriff freigeben. Wird auf dem Server ein anderes Betriebssystem ausgeführt (Ihr Computer läuft beispielsweise unter Windows und der Server unter UNIX), werden Sie im Modus für verteilte Analysen wahrscheinlich keinen Zugriff auf lokale Datendateien haben, selbst wenn sie sich in freigegebenen Ordnern befinden.

Der Modus für verteilte Analysen entspricht nicht dem einfachen Zugriff auf Datendateien, die sich auf einem anderen Computer im Netzwerk befinden. Der Zugriff auf Datendateien, die auf einem anderen Gerät im Netzwerk gespeichert sind, kann sowohl im Modus für lokale Analysen als auch im Modus für verteilte Analysen stattfinden. Im lokalen Modus können Sie von Ihrem lokalen Computer aus auf andere Geräte im Netzwerk zugreifen. Im verteilten Modus können Sie vom Remote-Server aus auf andere Geräte im Netz zugreifen.

Wenn Sie nicht sicher sind, ob Sie im Modus für lokale Analysen oder im Modus für verteilte Analysen arbeiten, schauen Sie in der Titelleiste eines der Dialogfelder für den Zugriff auf die Daten nach. Wenn der Titel des Dialogfelds das Wort *Entfernt* (wie beispielsweise Entfernte Datei öffnen) enthält oder der Text Remote-Server:

[Servername] im oberen Teil des Dialogfelds angezeigt wird, arbeiten Sie im Modus für verteilte Analysen.

Anmerkung: Dies gilt nur für Dialogfelder für den Zugriff auf Datendateien (beispielsweise zum Öffnen und Speichern von Daten, zum Öffnen von Datenbanken und zum Zuweisen des Datenlexikons). Bei allen anderen Dateitypen (beispielsweise Viewer-Dateien, Syntaxdateien und Skriptdateien) werden jeweils die lokal gespeicherten Dateien gezeigt.

So setzen Sie die Berechtigungen für freigegebene Laufwerke und Ordner:

- Klicken Sie im Fenster Arbeitsplatz auf den Ordner (das Verzeichnis) oder das Laufwerk, der/das freigegeben werden soll.
- Wählen Sie im Menü "Datei" den Befehl Eigenschaften aus.
- Klicken Sie auf die Registerkarte Freigabe, und klicken Sie dann auf Freigegeben als.

Weitere Informationen zum Freigeben von Laufwerken und Ordnern finden Sie in der Hilfe Ihres Betriebssystems.

Verfügbarkeit von Prozeduren im Modus für verteilte Analysen Modalwert

Im Modus für verteilte Analysen können Prozeduren nur verwendet werden, wenn diese sowohl auf dem lokalen Computer als auch auf dem Remote-Server installiert sind.

Wenn Sie optionale Komponenten lokal installiert haben, die auf dem Remote-Server nicht zur Verfügung stehen, und Sie zwischen dem lokalen Computer und dem Remote-Server wechseln, werden die entsprechenden Prozeduren aus dem Menü entfernt, und die Befehlssyntax wird lediglich zu Fehlern führen. Durch einen Wechsel zurück in den lokalen Modus werden die betroffenen Prozeduren wiederhergestellt.

Arbeiten mit UNC-Pfadangaben

Bei der SPSS-Version für Windows NT Server sind relative Pfadangaben im Modus für verteilte Analysen als relativ zum aktuellen Server zu verstehen, nicht als relativ zu Ihrem lokalen Computer. Eine Pfadangabe (wie *C:\Eigene Dateien\Marktanalyse.sav*) verweist nicht auf ein Verzeichnis und eine Datei auf Ihrem lokalen Laufwerk C:\,

sondern auf ein Verzeichnis und eine Datei auf der Festplatte des Remote-Servers. Wenn weder das Verzeichnis noch die Datei auf dem Remote-Server vorhanden sind, führt dies zu einer Fehlerausgabe der Befehlssyntax wie im folgenden Beispiel:

```
GET FILE='c:\mydocs\mydata.sav'.
```

Wenn Sie die SPSS-Version für Windows NT Server verwenden, können Sie beim Zugriff auf Datendateien mit Befehlssyntax UNC-Pfadangaben (UNC = Universal Naming Convention, Universelle Namenskonvention) verwenden. UNC-Pfadangaben weisen die folgende allgemeine Form auf:

```
\\Servername\Freigabe\Pfad\Dateiname
```

- *Servername* ist der Name des Computers, auf dem die Datendatei gespeichert ist.
- *Freigabe* ist der Ordner (das Verzeichnis) auf diesem Computer, der (oder das) freigegeben ist.
- *Pfad* sind die dem freigegebenen Verzeichnis untergeordneten Ordner (Unterordner bzw. Unterverzeichnisse).
- *Dateiname*ist der Name der Datendatei.

Ein Beispiel lautet folgendermaßen:

```
GET FILE='\\hqdev001\public\july\sales.sav'.
```

Wenn dem Computer kein Name zugewiesen wurde, können Sie seine IP-Adresse verwenden, wie in folgendem Beispiel:

```
GET FILE='\\204.125.125.53\public\july\sales.sav'.
```

Auch mit UNC-Pfadangaben können Sie nur auf Datendateien auf Geräten oder in Ordnern zugreifen, die ausdrücklich freigegeben wurden. Beim Modus für verteilte Analysen gilt dies auch für die Datendateien auf Ihrem lokalen Computer.

UNIX-Server. Auf UNIX-Plattformen gibt es kein Äquivalent zu UNC-Pfaden. Alle Verzeichnispfade müssen absolute Pfade sein, die beim Stamm des Servers beginnen; relative Pfade sind unzulässig. Wenn die Datendatei beispielsweise unter */bin/spss/data* gespeichert ist und das aktuelle Verzeichnis ebenfalls */bin/spss/data* ist, dann ist `GET FILE='sales.sav'` unzulässig; Sie müssen den gesamten Pfad angeben:

```
GET FILE='/bin/spss/data/sales.sav'.
```

Kapitel
5

Daten-Editor

Der Daten-Editor bietet eine praktische Methode zum Erstellen und Bearbeiten von SPSS-Datendateien, die der von Tabellenkalkulationen bekannten Methode ähnelt. Das Fenster des Daten-Editors wird automatisch geöffnet, wenn Sie eine SPSS-Sitzung beginnen.

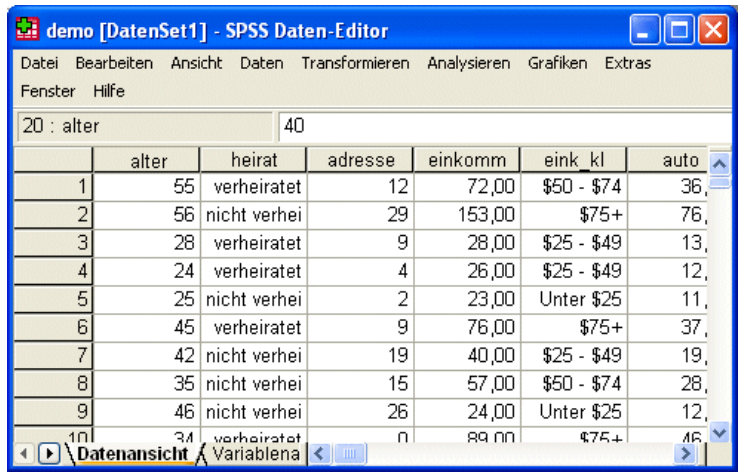
Der Daten-Editor stellt zwei Ansichten der Daten bereit:

- **Datenansicht.** In dieser Ansicht werden die eigentlichen Datenwerte oder die definierten Wertelabels angezeigt.
- **Variablenansicht.** In dieser Ansicht werden Informationen zu den Variablendefinitionen angezeigt. Dies umfasst die Variablen- und Wertelabels, den Datentyp (beispielsweise String, Datum oder numerisch), das Messniveau (nominal, ordinal oder metrisch) sowie benutzerdefinierte fehlende Werte.

In beiden Ansichten können Sie der Datendatei neue Informationen hinzufügen oder vorhandene Informationen bearbeiten und löschen.

Datenansicht

Abbildung 5-1
Datenansicht


demo [DatenSet1] - SPSS Daten-Editor
Datei Bearbeiten Ansicht Daten Transformieren Analysieren Grafiken Extras
Fenster Hilfe
20 : alter
40
alter heirat adresse einkomm eink_kl auto
1 55 verheiratet 12 72,00 $50 - $74 36,
2 56 nicht verhei 29 153,00 $75+ 76,
3 28 verheiratet 9 28,00 $25 - $49 13,
4 24 verheiratet 4 26,00 $25 - $49 12,
5 25 nicht verhei 2 23,00 Unter $25 11,
6 45 verheiratet 9 76,00 $75+ 37,
7 42 nicht verhei 19 40,00 $25 - $49 19,
8 35 nicht verhei 15 57,00 $50 - $74 28,
9 46 nicht verhei 26 24,00 Unter $25 12,
Datenansicht
Variablena

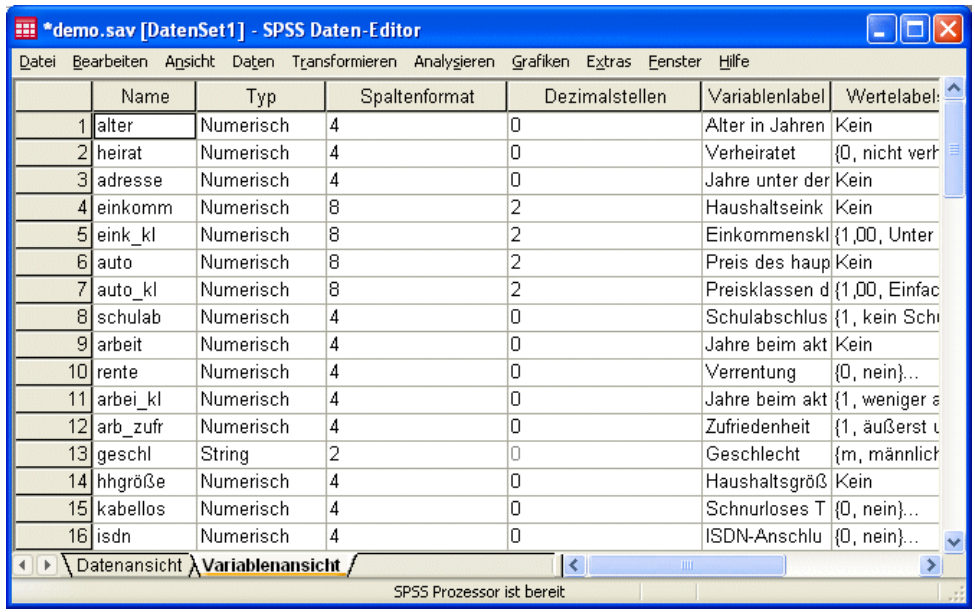
Viele Funktionen der Datenansicht ähneln den Funktionen von Anwendungen für die Tabellenkalkulation. Es gibt allerdings mehrere wichtige Unterschiede:

- Zeilen sind Fälle. Jede Zeile stellt einen Fall oder eine Beobachtung dar. So ist zum Beispiel jede Person, die einen Fragebogen ausfüllt, ein Fall.
- Spalten sind Variablen. Jede Spalte stellt eine Variable oder eine Eigenschaft dar, die gemessen wurde. Jedes Objekt auf einem Fragebogen ist zum Beispiel eine Variable.
- Zellen enthalten Werte. Jede Zelle enthält einen einzelnen Wert einer Variablen für einen Fall. Die Zelle befindet sich an der Schnittstelle von Fall und Variable. Zellen enthalten nur Datenwerte. Im Gegensatz zu Programmen für die Tabellenkalkulation können Zellen im Daten-Editor keine Formeln enthalten.
- In einer Datendatei enthalten alle Zeilen die gleiche Anzahl Zellen. Die Dimensionen der Datendatei werden von der Anzahl der Fälle und Variablen bestimmt. In alle Zellen können Daten eingegeben werden. Wenn Sie Daten in eine Zelle außerhalb der Grenzen der definierten Datendatei eingeben, erweitert SPSS das Datenfeld, so daß es alle Zeilen und/oder Spalten einschließt, die zwischen dieser Zelle und den Grenzen der Datendatei liegen. Innerhalb der Grenzen der Datendatei gibt es keine “leeren” Zellen. Bei numerischen Variablen

werden leere Zellen zum systemdefinierten fehlenden Wert konvertiert. Bei String-Variablen gelten leere Felder als gültiger Wert.

Variablenansicht

Abbildung 5-2
Variablenansicht


*demo.sav [DatenSet1] - SPSS Daten-Editor
Datei Bearbeiten Ansicht Daten Transformieren Analysieren Grafiken Extras Fenster Hilfe
Name Typ Spaltenformat Dezimalstellen Variablenlabel Wertelabel
1 alter Numerisch 4 0 Alter in Jahren Kein
2 heirat Numerisch 4 0 Verheiratet {0, nicht verh
3 adresse Numerisch 4 0 Jahre unter der Kein
4 einkomm Numerisch 8 2 Haushaltseink Kein
5 eink_kl Numerisch 8 2 Einkommenskl {1,00, Unter
6 auto Numerisch 8 2 Preis des haup Kein
7 auto_kl Numerisch 8 2 Preisklassen d {1,00, Einfac
8 schulab Numerisch 4 0 Schulabschlus {1, kein Schu
9 arbeit Numerisch 4 0 Jahre beim akt Kein
10 rente Numerisch 4 0 Verrentung {0, nein}...
11 arbei_kl Numerisch 4 0 Jahre beim akt {1, weniger a
12 arb_zufr Numerisch 4 0 Zufriedenheit {1, äußerst u
13 geschl String 2 0 Geschlecht {m, männlich
14 hhgröße Numerisch 4 0 Haushaltsgröß Kein
15 kabellos Numerisch 4 0 Schnurloses T {0, nein}...
16 isdn Numerisch 4 0 ISDN-Anschlu {0, nein}...
Datenansicht Variablenansicht
SPSS Prozessor ist bereit

In der Variablenansicht werden die Attribute alle Variablen in der Datendatei angezeigt. Die Variablenansicht ist wie folgt aufgeteilt:

- Die Zeilen stellen Variablen dar.
- Die Spalten stellen die Attribute der Variablen dar.

Sie können Variablen hinzufügen und löschen, und Sie können die folgenden Variablenattribute ändern:

- Variablenname
- Datentyp
- Anzahl Ziffern oder Zeichen
- Anzahl Dezimalstellen

- Beschreibende Variablen- und Wertelabels
- Benutzerdefinierte fehlende Werte
- Spaltenbreite
- Meßniveau

Alle diese Attribute werden beim Speichern der Datendatei gespeichert.

Neben den Methoden, mit denen Variableneigenschaften in der Variablenansicht definiert werden, gibt es zwei weitere Methoden zum Definieren von Variableneigenschaften:

- Mit dem Assistenten zum Kopieren von Dateneigenschaften können Sie eine externe SPSS-Datendatei oder ein anderes Daten-Set, das in der aktuellen Sitzung verfügbar ist, als Vorlage für die Definition von Datei- und Variableneigenschaften in der Arbeitsdatei verwenden. Sie können außerdem Variablen in der Arbeitsdatei als Vorlagen für andere Variablen in der Arbeitsdatei verwenden. Sie können den Assistenten zum Kopieren von Dateneigenschaften starten, indem Sie im Fenster des Daten-Editors im Menü “Daten” den Befehl “Dateneigenschaften kopieren” auswählen.
- Mit dem Befehl “Variableneigenschaften definieren” (ebenfalls im Menü “Daten” des Daten-Editors) können Sie Ihre Daten durchsuchen und eine Liste mit allen eindeutigen Datenwerten für die ausgewählten Variablen erstellen, Werte ohne Labels ausfindig machen und Werte automatisch mit Labels versehen. Diese Methode ist insbesondere für kategoriale Variablen sinnvoll, bei denen numerische Codes Kategorien darstellen, beispielsweise 0 = *männlich* und 1 = *weiblich*.

So zeigen Sie die Attribute von Variablen an und legen diese fest

▶ Aktivieren Sie das Fenster des Daten-Editors.

▶ Doppelklicken Sie in der Datenansicht auf den Namen einer Variablen im Kopf einer Spalte, oder klicken Sie auf die Registerkarte Variablenansicht.

▶ Wenn Sie eine neue Variable definieren möchten, geben Sie einen Namen in eine beliebige leere Zeile ein.

▶ Wählen Sie die Attribute aus, die Sie festlegen oder ändern möchten.

Variablennamen

Beim Benennen von Variablen gelten die folgenden Regeln:

- Namen müssen mit einem Buchstaben beginnen. Für die übrigen Zeichen können beliebige Buchstaben oder Ziffern, Punkte sowie die Symbole @, #, _ und $ verwendet werden.
- Das letzte Zeichen eines Variablennamens darf kein Punkt sein.
- Es empfiehlt sich, keine Variablennamen zu verwenden, die mit einem Unterstrich enden. Dadurch vermeiden Sie Konflikte mit Variablen, die von Prozeduren angelegt werden.
- Die Namen dürfen nicht länger als 64 Byte sein. 64 Byte entsprechen in Single-Byte-Sprachen (z. B. Englisch, Französisch, Deutsch, Spanisch, Italienisch, Hebräisch, Russisch, Griechisch, Arabisch, Thai) normalerweise 64 Zeichen und in Double-Byte-Sprachen (z. B. Japanisch, Chinesisch, Koreanisch) normalerweise 32 Zeichen.
- Leer- und Sonderzeichen, wie beispielsweise !, ?, ' und *, dürfen nicht verwendet werden.
- Variablennamen müssen eindeutig sein. Doppelt vorkommende Namen sind nicht zulässig.
- Reservierte Schlüsselwörter können nicht als Variablennamen verwendet werden. Reservierte Schlüsselwörter sind: `ALL`, `AND`, `BY`, `EQ`, `GE`, `GT`, `LE`, `LT`, `NE`, `NOT`, `OR`, `TO` und `WITH`.
- Variablennamen können aus einer beliebigen Kombination aus Klein- und Großbuchstaben bestehen. Die Groß- und Kleinschreibung bleibt auch bei der Anzeige erhalten.
- Wenn lange Variablennamen in der Ausgabe mehrere Zeilen einnehmen, versucht SPSS den Zeilenumbruch bei Unterstrichen, Punkten und dem Wechsel von Klein- zu Großschreibung durchzuführen.

Meßniveau einer Variablen

Das Meßniveau kann als Skala (für numerische Daten in Form einer Intervall- oder Verhältnisskala), ordinal oder nominal angegeben werden. Nominale und ordinale Daten können entweder aus einem String (alphanumerisch) oder Zahlen bestehen.

- **nominal.** Eine Variable kann als nominal behandelt werden, wenn ihre Kategorien sich nicht in eine natürliche Reihenfolge bringen lassen, z. B. die Unternehmensabteilung, in der eine Person arbeitet.
- **ordinal.** Eine Variable kann als ordinal behandelt werden, wenn ihre Kategorien sich in eine natürliche Reihenfolge bringen lassen, z. B. Zufriedenheit mit Kategorien von sehr unzufrieden bis sehr zufrieden.
- **metrisch.** Eine Variable kann als metrisch behandelt werden, wenn man sinnvolle Aussagen über die Abstände zwischen den Werten machen kann. Metrische Variablen sind z. B. Alter (in Jahren) oder Einkommen (in Geldeinheiten).

Anmerkung: Bei ordinalen String-Variablen wird angenommen, daß die Reihenfolge der Kategorien der alphabetischen Reihenfolge der String-Werte entspricht. Bei einer String-Variablen mit den Werten *Schwach*, *Mittel* und *Stark* werden die Kategorien beispielsweise in der Reihenfolge *Mittel*, *Schwach*, *Stark* und somit falsch angeordnet. Im allgemeinen ist die Verwendung von numerischem Code für ordinale Daten günstiger.

Die Standardeinstellung für das Messniveau wird für neue Variablen, die in einer Sitzung erstellt wurde, für Daten, die aus Dateien in einem externen Dateiformat stammen, und für SPSS-Datendateien, die mit Programmversionen vor 8.0 erstellt wurden, anhand folgender Regeln vorgenommen:

- Numerische Variablen mit weniger als 24 eindeutigen Werten und String-Variablen werden als nominal festgelegt.
- Numerische Variablen mit 24 oder mehr eindeutigen Werten werden als metrisch festgelegt.

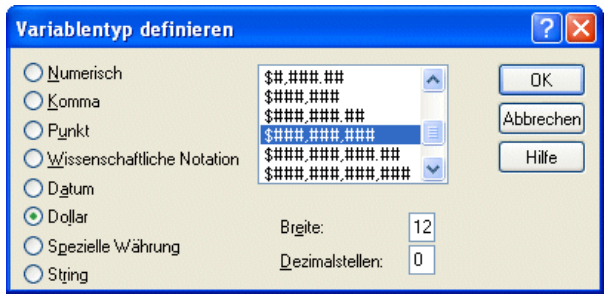
Im Dialogfeld “Optionen” können Sie den Wert ändern, anhand dessen bei numerischen Variablen zwischen metrisch und nominal unterschieden wird. Für weitere Informationen siehe “Optionen: Interaktiv” in Kapitel 44 auf S. 678.

Mithilfe des Dialogfelds “Variableneigenschaften definieren”, das in dem Menü “Daten” verfügbar ist, können Sie das richtige Messniveau zuweisen. Für weitere Informationen siehe “Zuweisen des Meßniveaus” in Kapitel 7 auf S. 131.

Variablentyp

Mit “Variablentyp definieren” wird für jede Variable der Datentyp angegeben. In der Standardeinstellung sind alle neuen Variablen als numerisch festgelegt. Mit “Variablentyp definieren” können Sie den Datentyp ändern. Dabei ist der Inhalt des Dialogfelds “Variablentyp definieren” von dem jeweils ausgewählten Datentyp abhängig. Bei einigen Datentypen gibt es Textfelder für die Breite und die Anzahl der Dezimalstellen, bei anderen Datentypen können Sie einfach ein Format aus einer Liste mit Beispielen auswählen.

Abbildung 5-3
Dialogfeld “Variablentyp definieren”



Die folgenden Datentypen sind verfügbar:

Numerisch. Eine Variable, deren Werte Zahlen sind. Die Werte werden im numerischen Standardformat angezeigt. Numerische Wert können im Daten-Editor im Standardformat oder in wissenschaftlicher Notation eingegeben werden.

Komma. Eine numerische Variable, deren Werte mit Kommata als Tausender-Trennzeichen und Punkt als Dezimaltrennzeichen angezeigt werden. Numerische Werte für Kommavariablen können im Daten-Editor mit oder ohne Kommata oder in wissenschaftlicher Notation eingegeben werden. Die Werte können rechts neben dem Dezimaltrennzeichen kein Komma enthalten.

Punkt. Eine numerische Variable, deren Werte mit Punkten als Tausender-Trennzeichen und Komma als Dezimaltrennzeichen angezeigt werden. Numerische Werte für Punktvariablen können im Daten-Editor mit oder ohne Punkte oder in wissenschaftlicher Notation eingegeben werden. Die Werte können rechts neben dem Dezimaltrennzeichen keinen Punkt enthalten.

Wissenschaftliche Notation. Eine numerische Variable, deren Werte mit einem E und einer Zehnerpotenz mit Vorzeichen angezeigt werden. Numerische Werte für diese Variablen können im Daten-Editor mit oder ohne Potenz eingegeben werden. Dem Exponenten kann entweder ein E oder ein D (mit oder ohne Vorzeichen) oder ein Vorzeichen allein vorangestellt werden, beispielsweise 123, 1,23E2, 1,23D2, 1,23E+2 oder 1,23+2.

Datum. Eine numerische Variable, deren Werte in einem der Datums- oder Uhrzeitformate angezeigt werden. Wählen Sie ein Format aus der Liste aus. Sie können Datumsangaben mit Schrägstrichen, Bindestrichen, Punkten, Kommata oder Leerzeichen als Trennzeichen eingeben. Bei zweistelligen Jahresangaben hängt das Jahrhundert von den Einstellungen unter "Optionen" ab (wählen Sie dazu im Menü "Bearbeiten" den Befehl Optionen aus, und klicken Sie dann auf die Registerkarte Daten).

Dollar. Eine numerische Variable mit führendem Dollarzeichen ($), deren Werte mit Kommata als Tausender-Trennzeichen und Punkt als Dezimaltrennzeichen angezeigt werden. Die Werte können mit und ohne das führende Dollarzeichen eingegeben werden.

Spezielle Währung. Eine numerische Variable, deren Werte in einem der Währungsformate angezeigt wird. Währungsformate werden im Dialogfeld "Optionen" auf der Registerkarte "Währung" definiert. Zeichen, die in einem Währungsformat festgelegt wurden, können nicht für die Dateneingabe genutzt werden. Die Zeichen werden jedoch im Daten-Editor angezeigt.

String. Eine Variable, deren Werte nicht numerisch sind und die daher nicht in den Berechnungen verwendet werden. Die Werte dürfen beliebige Zeichen bis zur festgelegten Höchstlänge enthalten. Groß- und Kleinbuchstaben werden als separate Buchstaben betrachtet. Dieser Typ ist auch als alphanumerische Variable bekannt.

So definieren Sie einen Variablentyp

- Klicken Sie auf die Schaltfläche in der Zelle *Typ* der Variablen, die Sie definieren möchten.
- Wählen Sie im Dialogfeld "Variablentyp definieren" den Datentyp aus.
- Klicken Sie auf OK.

Der Unterschied zwischen Eingabe- und Anzeigeformaten

Je nach Format kann die Anzeige von Werten in der Datenansicht von den eingegebenen und tatsächlich intern gespeicherten Werten abweichen. Im folgenden finden Sie einige allgemeine Richtlinien:

- In den Formaten numerisch, Komma und Punkt können Sie Werte mit jeder beliebigen Anzahl (bis zu 16) Dezimalstellen eingeben. Der gesamte Wert wird intern gespeichert. In der Datenansicht wird nur die definierte Anzahl der Stellen angezeigt. Werte mit mehr Dezimalstellen werden gerundet. In Berechnungen wird allerdings immer der vollständige Wert verwendet.
- Bei String-Variablen werden alle Werte bis zur maximalen Länge rechts mit Leerzeichen aufgefüllt. Bei einer String-Variablen mit einer maximalen Breite von 3 wird der Wert *Nein* intern als `'Nein '` gespeichert und ist somit nicht das gleiche wie `' Nein'`.
- Bei Datumsformaten können Sie Schrägstriche, Bindestriche, Leerzeichen, Kommata oder Punkte als Trennzeichen zwischen den Werten für Tag, Monat und Jahr verwenden. Für die Monatswerte können Sie Ziffern, Abkürzungen von drei Buchstaben Länge oder vollständige Namen eingeben. Datumsangaben im allgemeinen Format “tt-mmm-jj” werden mit Bindestrichen als Trennzeichen und Abkürzungen von drei Buchstaben Länge für den Monat eingegeben. Datumsangaben im allgemeinen Format “tt-mm-jj” und “mm/tt/jj” werden mit Schrägstrichen als Trennzeichen und Zahlen für den Monat eingegeben. Die Daten werden intern als Anzahl der Sekunden gespeichert, die seit dem 14. Oktober 1582 vergangen sind. Den Jahrhundertbereich für zweistellige Jahresangaben können Sie in den Einstellungen unter “Optionen” angeben. Wählen Sie dazu im Menü “Bearbeiten” den Befehl Optionen und anschließend die Registerkarte Daten aus.
- In Zeitformaten können Sie Doppelpunkte, Punkte oder Leerzeichen als Trennzeichen zwischen Stunden, Minuten und Sekunden verwenden. Zeiten werden mit Doppelpunkten als Trennzeichen angezeigt. Datumsangaben werden intern als die Anzahl von Sekunden gespeichert, die seit dem 14. Oktober 1582 verstrichen sind.

Variablenlabels

Sie können aussagekräftige Variablenlabels bis zu 256 Zeichen Länge (128 Zeichen für Double-Byte-Sprachen) zuweisen. Variablenlabels können Leerzeichen und reservierte Zeichen enthalten, die in Variablennamen nicht zulässig sind.

So legen Sie Variablenlabels fest

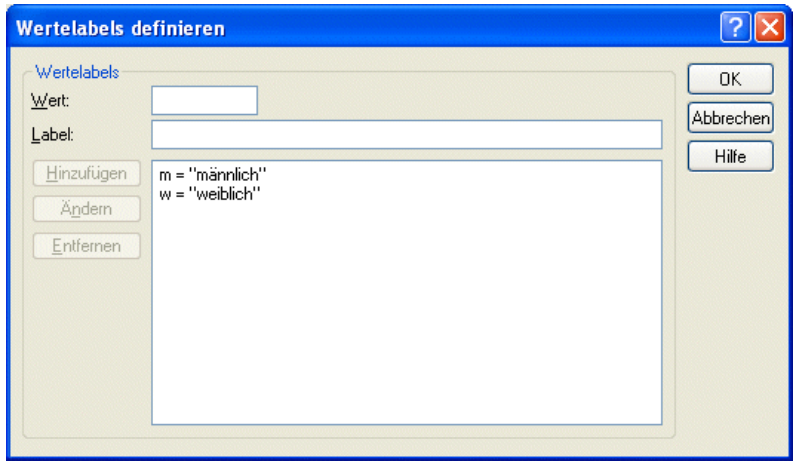
- ▶ Aktivieren Sie das Fenster des Daten-Editors.
- ▶ Doppelklicken Sie in der Datenansicht auf den Namen einer Variablen im Kopf einer Spalte, oder klicken Sie auf die Registerkarte Variablenansicht.
- ▶ Geben Sie in der Zelle *Label* für die Variable ein aussagekräftiges Variablenlabel ein.

Wertelabels

Sie können jedem Wert einer Variable ein beschreibendes Wertelabel zuordnen. Dies ist besonders nützlich, wenn Ihre Datendatei numerische Codes zur Darstellung nichtnumerischer Kategorien verwendet (zum Beispiel die Codes 1 und 2 für *Männlich* und *Weiblich*).

- ■ Wertelabels können bis zu 120 Byte umfassen.
- ■ Für lange String-Variablen (String-Variablen mit mehr als 8 Zeichen Länge) sind keine Wertelabels verfügbar.

Abbildung 5-4
Dialogfeld "Wertelabels"



So legen Sie Wertelabels fest

- Klicken Sie auf die Schaltfläche in der Zelle *Werte* der Variablen, die Sie definieren möchten.
- Geben Sie für jeden Wert den Wert und ein Label ein.
- Klicken Sie auf Hinzufügen, um das Wertelabel einzugeben.
- Klicken Sie auf OK.

Einfügen von Zeilenumbrüchen in Labels

Bei Variablenlabels und Wertelables werden in Pivot-Tabellen und Diagrammen automatisch Zeilenumbrüche eingefügt, wenn die Zelle bzw. der Bereich nicht breit genug ist, um das gesamte Label in einer Zeile anzuzeigen. Außerdem können Sie die Ergebnisse bearbeiten, um manuelle Zeilenumbrüche einzufügen, wenn der Label-Text an einer anderen Stelle umbrechen soll. Des Weiteren können Sie

Variablenlabels und Wertelables erstellen, bei denen der Text *immer* an festgelegten Punkten umbricht und auf mehrere Zeilen verteilt angezeigt wird:

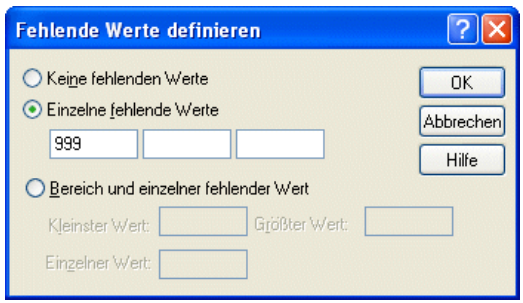
- Bei Variablenlabels wählen Sie in der Variablenansicht des Daten-Editors die Zelle Label für die Variable aus.
- Bei Wertelabels wählen Sie in der Variablenansicht des Daten-Editors die Zelle Werte für die Variable aus, klicken Sie auf die Schaltfläche in der Zelle, und wählen Sie im Dialogfeld "Wertelabels definieren" das zu ändernde Label aus.
- Geben Sie an der Stelle, an der der Zeilenumbruch erfolgen soll, \n ein.

Das \n wird in den Pivot-Tabellen bzw. Diagrammen nicht angezeigt; es wird als Zeichen für den Zeilenumbruch interpretiert.

Fehlende Werte

Mit der Option "Fehlende Werte" werden bestimmte Datenwerte zu **benutzerdefinierten fehlenden Werten** deklariert. So ist es zum Beispiel sinnvoll, zu unterscheiden, ob Daten fehlen, weil eine befragte Person die Auskunft verweigerte oder weil die Frage sich nicht auf die befragte Person bezog. Datenwerte, die als benutzerdefinierte fehlende Werte angegeben sind, werden zur Sonderbehandlung gekennzeichnet und von den meisten Berechnungen ausgeschlossen.

Abbildung 5-5
Dialogfeld "Fehlende Werte"



- Sie können entweder bis zu drei diskrete (einzelne) fehlende Werte, einen Bereich fehlender Werte oder einen Bereich und einen diskreten Wert eingeben.

- Bereiche können nur bei numerischen Variablen angegeben werden.
- Fehlende Werte können nicht für lange String-Variablen (String-Variablen mit mehr als 8 Zeichen Länge) definiert werden.

Fehlende Werte für String-Variablen. Zunächst werden alle String-Variablen, einschließlich der Werte "Null" und "Leer", als gültig betrachtet, sofern diese nicht explizit als "fehlend" definiert worden sind. Wenn Sie String-Variablen mit den Werten "Null" oder "Leer" als fehlend definieren möchten, geben Sie in eines der Felder von Einzelne fehlende Werte ein einfaches Leerzeichen ein.

So definieren Sie fehlende Werte:

▶ Klicken Sie auf die Schaltfläche in der Zelle *Fehlende Werte* der Variablen, die Sie definieren möchten.

▶ Geben Sie die Werte oder den Bereich der Werte ein, welche die fehlenden Daten repräsentieren.

Zunächst werden alle String-Variablen, einschließlich der Werte "Null" und "Leer", als gültig betrachtet, sofern diese nicht explizit als "fehlend" definiert worden sind. Wenn Sie String-Variablen mit den Werten "Null" oder "Leer" als fehlend definieren möchten, geben Sie in eines der Felder von Einzelne fehlende Werte ein einfaches Leerzeichen ein.

Spaltenbreite

Sie können die Spaltenbreite als Anzahl der angezeigten Zeichen festlegen. Die Spaltenbreite kann auch in der Datenansicht geändert werden, indem Sie auf eine Spaltenbegrenzung klicken und sie an die gewünschte Stelle ziehen.

Spaltenformate wirken sich nur auf die Anzeige der Werte im Daten-Editor aus. Änderungen der Spaltenbreite ändern nicht die definierte Länge einer Variablen. Wenn die definierte und tatsächliche Länge eines Werts größer ist als die Breite der Spalte, werden in der Datenansicht Sternchen (*) angezeigt.

Variablenausrichtung

Mit der Ausrichtung wird die Anzeige von Datenwerten und/oder Wertelabels in der Datenansicht festgelegt. Numerische Variablen werden in der Standardeinstellung rechts, String-Variablen links ausgerichtet. Diese Einstellung gilt nur für die Anzeige in der Datenansicht.

Zuweisen von Variablenattributen zu mehreren Variablen

Nachdem Sie Attribute zur Variablendefinition festgelegt haben, können Sie ein oder mehrere Attribute kopieren und einer oder mehreren Variablen zuweisen.

Das Zuweisen der Variablenattribute erfolgt über einfaches Kopieren und Einfügen. Sie verfügen über folgende Möglichkeiten:

- Einzelne Attribute (beispielsweise Wertelabels) können kopiert und in die gleiche Attributzelle einer oder mehrerer Variablen eingefügt werden.
- Alle Attribute einer Variablen können kopiert und in eine oder mehrere andere Variablen eingefügt werden.
- Mehrere neue Variablen mit sämtlichen Attributen einer kopierten Variable können erstellt werden.

Zuweisen von Variablenattributen zu anderen Variablen

So übertragen Sie einzelne Attribute aus einer bereits definierten Variablen:

▶ Wählen Sie in der Variablenansicht die Attributzelle aus, die Sie auf andere Variablen übertragen möchten.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Bearbeiten
Kopieren

▶ Wählen Sie die Attributzelle(n) aus, in die Sie das Attribut übertragen möchten. (Es können mehrere Zielvariablen ausgewählt werden.)

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Bearbeiten
Einfügen

Wenn Sie die Attribute in leere Zeilen einfügen, werden neue Variablen erstellt, wobei allen Attributen mit Ausnahme des ausgewählten Attributs Standardwerte zugewiesen werden.

So übertragen Sie alle Attribute aus einer bereits definierten Variable:

- Wählen Sie in der Variablenansicht die Zeilennummer der Variablen aus, deren Attribute Sie übertragen möchten. (Die gesamte Zeile wird markiert.)
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Kopieren
- Klicken Sie auf die Zeilennummer(n) der Variablen, der/denen Sie die Attribute zuweisen möchten. (Es können mehrere Zielvariablen ausgewählt werden.)
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Einfügen

Erstellen von mehreren neuen Variablen mit übereinstimmenden Attributen

- Klicken Sie in der Variablenansicht auf die Zeilennummer der Variablen, deren Attribute Sie auf die neue Variable übertragen möchten. (Die gesamte Zeile wird markiert.)
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Kopieren
- Klicken Sie auf die Nummer der leeren Zeile unterhalb der letzten definierten Variablen in der Datendatei.
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Variablen einfügen...
- Geben Sie im Dialogfeld “Variablen einfügen” die Anzahl der Variablen ein, die Sie erstellen möchten.
- Geben Sie ein Präfix und eine Anfangsnummer für die neuen Variablen ein.
- Klicken Sie auf OK.

Die Namen der neuen Variablen bestehen aus dem angegebenen Präfix und einer laufenden Nummer (ab der angegebenen Anfangsnummer).

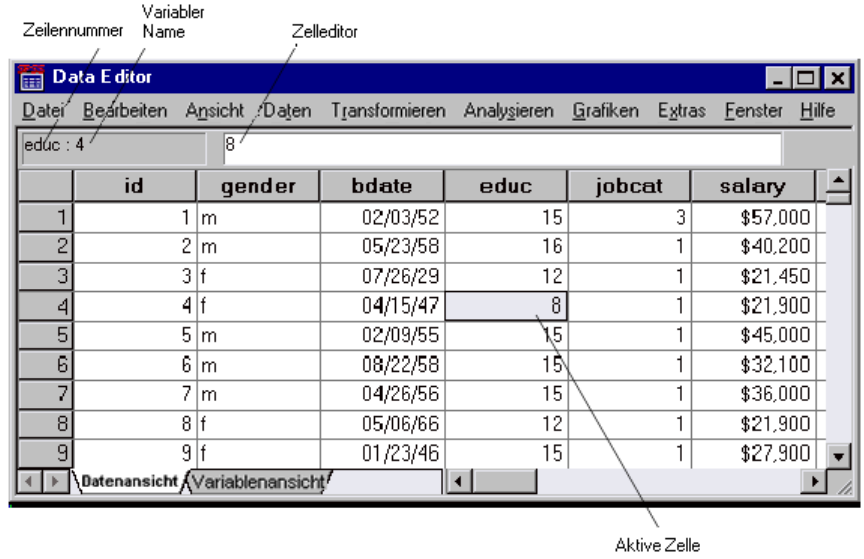
Eingeben von Daten

In der Datenansicht können Sie die Daten direkt in den Daten-Editor eingeben. Sie können Daten in beliebiger Reihenfolge eingeben. Sie können Daten nach Fall oder Variable, für ausgewählte Bereiche oder einzelne Zellen eingeben.

- Die aktive Zelle wird hervorgehoben.
- In der linken oberen Ecke des Daten-Editors werden der Name der Variablen und die Zeilennummer der aktiven Zelle angezeigt.
- Wenn Sie eine Zelle auswählen und einen Datenwert eingeben, wird der Wert im Zellen-Editor am oberen Rand des Daten-Editors angezeigt.
- Datenwerte werden nicht aufgezeichnet, bis Sie die Eingabetaste drücken oder eine andere Zelle wählen.
- Wenn Sie andere Daten als einfache numerische Daten eingeben möchten, müssen Sie zuerst den Variablentyp definieren.

Wenn Sie einen Wert in eine leere Spalte eingeben, wird vom Daten-Editor automatisch eine neue Variable erstellt und ein neuer Variablenname zugewiesen.

Abbildung 5-6
Arbeitsdatei in der Datenansicht


Zeilennummer
Variabler Name
Zelleditor
Data Editor
Datei Bearbeiten Ansicht Daten Transformieren Analysieren Grafiken Extras Fenster Hilfe
educ : 4
8
id gender bdate educ jobcat salary
1 1 m 02/03/52 15 3 $57,000
2 2 m 05/23/58 16 1 $40,200
3 3 f 07/26/29 12 1 $21,450
4 4 f 04/15/47 8 1 $21,900
5 5 m 02/09/55 15 1 $45,000
6 6 m 08/22/58 15 1 $32,100
7 7 m 04/26/56 15 1 $36,000
8 8 f 05/06/66 12 1 $21,900
9 9 f 01/23/46 15 1 $27,900
Datenansicht
Variablenansicht
Aktive Zelle

So geben Sie numerische Daten ein:

- Wählen Sie in der Datenansicht eine Zelle aus.
- Geben Sie den Datenwert ein. (Der Wert wird im Zellen-Editor am oberen Rand des Daten-Editors angezeigt.)
- Drücken Sie die Eingabetaste, oder wählen Sie eine andere Zelle aus, um den Wert aufzuzeichnen.

So geben Sie nichtnumerische Daten ein

- Doppelklicken Sie in der Datenansicht auf den Namen einer Variablen im Kopf einer Spalte, oder klicken Sie auf die Registerkarte Variablenansicht.
- Klicken Sie auf die Schaltfläche in der Zelle *Typ* der Variablen.
- Wählen Sie im Dialogfeld “Variablentyp definieren” den Datentyp aus.
- Klicken Sie auf OK.

- Doppelklicken Sie auf die Zeilennummer, oder klicken Sie auf die Registerkarte Datenansicht.
- Geben Sie die Daten in die Spalte für die neu definierte Variable ein.

So verwenden Sie Wertelabels bei der Dateneingabe:

- Falls in der Datenansicht gegenwärtig keine Wertelabels angezeigt werden, wählen Sie die folgenden Befehle aus den Menüs aus:

  Ansicht
  Wertelabels
- Klicken Sie auf die Zelle, in der Sie den Wert eingeben möchten.
- Wählen Sie aus der Dropdown-Liste ein Wertelabel aus.

  Der Wert wird eingegeben und das Wertelabel in der Zelle angezeigt.

  Anmerkung: Dies ist jedoch nur möglich, wenn für die Variable Wertelabels definiert wurden.

Einschränkungen für die Datenwerte im Daten-Editor

Der definierte Variablentyp und die definierte Variablenlänge bestimmen, welche Werte in der Datenansicht in die Zelle eingegeben werden können.

- Wenn Sie ein für den definierten Variablentyp nicht zugelassenes Zeichen eingeben, gibt der Daten-Editor ein akustisches Signal aus und nimmt die Eingabe nicht an.
- Bei String-Variablen sind nicht mehr Zeichen erlaubt, als die definierte Länge zuläßt.
- Bei numerischen Variablen können ganzzahlige Werte eingegeben werden, welche die definierte Länge überschreiten, aber der Daten-Editor zeigt entweder die Werte in wissenschaftlicher Notation oder Sterne in der Zelle an, um zu kennzeichnen, daß der Wert länger als zulässig ist. Um den Wert in der Zelle anzuzeigen, müssen Sie die definierte Länge der Variablen ändern. (*Anmerkung*: Das Ändern der Spaltenbreite hat keinen Einfluß auf die Länge der Variablen.)

Bearbeiten von Daten

Mit dem Daten-Editor können Sie die Datenwerte in der Datenansicht auf verschiedene Arten bearbeiten. Sie verfügen über folgende Möglichkeiten:

- Datenwerte ändern
- Datenwerte ausschneiden, kopieren und einfügen
- Fälle hinzufügen und löschen
- Variablen hinzufügen und löschen
- Reihenfolge der Variablen ändern

Ersetzen oder Ändern eines Datenwerts

So löschen Sie den alten Wert und geben einen neuen Wert ein:

- Doppelklicken Sie in der Datenansicht auf die Zelle. (Der Wert der Zelle wird im Zellen-Editor angezeigt.)
- Bearbeiten Sie den Wert direkt in der Zelle oder im Zellen-Editor.
- Drücken Sie die Eingabetaste, oder setzen Sie den Mauszeiger in eine andere Zelle, um den neuen Wert aufzuzeichnen.

Ausschneiden, Kopieren und Einfügen von Datenwerten

Im Daten-Editor können Sie einzelne Werte aus Zellen oder Gruppen von Werten ausschneiden, kopieren und einfügen. Sie verfügen über folgende Möglichkeiten:

- Einen einzelnen Zellenwert in eine andere Zelle verschieben oder kopieren.
- Einen einzelnen Zellenwert in eine Gruppe von Zellen verschieben oder kopieren.
- Die Werte für einen einzelnen Fall (Zeile) in mehrere Fälle verschieben oder kopieren.
- Die Werte für eine einzelne Variable (Spalte) in mehrere Variablen verschieben oder kopieren.
- Eine Gruppe von Zellenwerten in eine andere Gruppe von Zellen verschieben oder kopieren.

Umwandlung für eingefügte Werte im Daten-Editor

Wenn die definierten Variablentypen der Quell- und Zielzellen nicht übereinstimmen, versucht der Daten-Editor, den Wert zu konvertieren. Wenn eine Umwandlung nicht möglich ist, wird der systemdefinierte fehlende Wert in die Zielzelle eingefügt.

Umwandlung von numerischen Formaten oder Datumsformaten in Strings. Numerische Formate (zum Beispiel numerisch, Dollar, Punkt oder Komma) und Datumsformate werden zu Strings konvertiert, wenn sie in eine Zelle für String-Variablen eingefügt werden. Der Stringwert ist der in der Zelle angezeigte numerische Wert. So wird zum Beispiel bei einer Variablen im Dollarformat das angezeigte Dollarzeichen zum Bestandteil des Stringwerts. Werte, welche die definierte Länge der String-Variablen übersteigen, werden abgeschnitten.

Umwandlung von Strings in numerische Werte oder Datumswerte. Stringwerte, die akzeptable Zeichen für das numerische Format oder das Datumsformat der Zielzelle enthalten, werden in den äquivalenten numerischen Wert oder Datumswert konvertiert. So wird zum Beispiel der String-Wert “25/12/91” in ein gültiges Datum konvertiert, falls das Format der Zielzelle vom Typ Tag-Monat-Jahr ist. Falls das Format der Zielzelle aber vom Typ Monat-Tag-Jahr ist, dann wird er in den systemdefinierten fehlenden Wert konvertiert.

Umwandlung von Datumswerten in numerische Werte. Werte für Datum und Uhrzeit werden in eine Anzahl von Sekunden umgewandelt, wenn die Zielzelle im numerischen Format ist (zum Beispiel numerisch, Dollar, Punkt oder Komma). Da Datumsangaben intern als die Anzahl der seit dem 14. Oktober 1582 vergangenen Sekunden gespeichert werden, kann das Umwandeln von Daten in numerische Werte zu extrem großen Zahlen führen. Das Datum 10/29/91 wird beispielsweise in den numerischen Wert 12.908.073.600 umgewandelt.

Umwandlung von numerischen Werten in Datums- oder Uhrzeitangaben. Numerische Werte werden in Datums- oder Uhrzeitangaben umgewandelt, wenn der Wert eine Anzahl von Sekunden darstellt, die in eine gültige Uhrzeit- oder Datumsangabe umgewandelt werden kann. Bei Datumsangaben werden numerische Werte unter 86.400 in den systemdefinierten fehlenden Wert umgewandelt.

Einfügen von neuen Fällen

Durch die Eingabe von Daten in eine Zelle in einer leeren Zeile wird automatisch ein neuer Fall angelegt. Für alle anderen Variablen dieses Falls fügt der Daten-Editor den systemdefinierten fehlenden Wert ein. Wenn sich zwischen dem neuen Fall und den bereits vorhandenen Fällen Leerzeilen befinden, dann werden die Leerzeilen ebenfalls neue Fälle mit dem systemdefinierten fehlenden Wert für alle Variablen. Sie können neue Fälle auch zwischen vorhandenen Fällen einfügen.

So fügen Sie einen neuen Fall zwischen vorhandenen Fällen ein:

- Wählen Sie in der Datenansicht eine Zelle in dem Fall (in der Zeile) unterhalb der Position aus, an der Sie den neuen Fall einfügen möchten.
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Fälle einfügen

  Für den Fall wird eine neue Zeile eingefügt, und alle Variablen erhalten den systemdefinierten fehlenden Wert.

Einfügen von neuen Variablen

Wenn Sie in der Datenansicht in eine leere Spalte oder in der Variablenansicht in eine leere Zeile Daten eingeben, wird automatisch eine neue Variable erstellt. Dieser Variable wird ein Standardname (Präfix *var* und eine laufende Nummer) und ein Standard-Datentyp (numerisch) zugewiesen. Der Daten-Editor fügt für alle Fälle der neuen Variable die systemdefinierten fehlenden Werte ein. Wenn in der Datenansicht leere Spalten oder in der Variablenansicht leere Zeilen zwischen der neuen Variablen und bereits vorhandenen Variablen stehen, werden diese Zeilen und Spalten ebenfalls in neue Variablen umgewandelt. Auch diesen Variablen wird der systemdefiniert fehlende Wert für alle Fälle zugewiesen. Sie können neue Variablen auch zwischen vorhandenen Variablen einfügen.

So fügen Sie eine neuen Variable zwischen vorhandenen Variablen ein:

- Wählen Sie eine beliebige Zelle rechts neben (in der Datenansicht) oder direkt unter (in der Variablenansicht) der Position aus, an der Sie die neue Variable einfügen möchten.

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Variable einfügen

  Neue Variablen werden mit dem systemdefinierten fehlenden Wert für alle Fälle eingefügt.

So verschieben Sie Variablen:

- Um die Variable auszuwählen, klicken Sie in der Datenansicht auf den Variablennamen oder in der Variablenansicht auf die Zeilennummer für die Variable.
- Ziehen Sie die Variable an die neue Position.
- Wenn Sie die Variable zwischen zwei vorhandenen Variablen einfügen möchten: Legen Sie in der Datenansicht die Variable in der Variablenspalte rechts neben der Stelle ab, an der Sie die Variable platzieren möchten, oder legen Sie die Variable in der Variablenansicht in der Variablenzeile unterhalb der Stelle ab, an der Sie die Variable platzieren möchten.

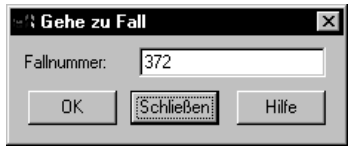
So ändern Sie den Datentyp:

Sie können den Datentyp einer Variablen jederzeit ändern, indem sind das Dialogfeld "Variablentyp definieren" in der Variablenansicht verwenden. Der Daten-Editor versucht, bereits vorhandene Werte in den neuen Typ zu konvertieren. Ist eine Umwandlung nicht möglich, wird der systemdefinierte fehlende Wert zugewiesen. Die Umwandlungsregeln sind dieselben wie beim Einfügen von Datenwerten in eine Variable mit anderem Formattyp. Falls die Änderung des Datenformats zum Verlust von Definitionen fehlender Werte oder von Wertelabels führen könnte, zeigt der Daten-Editor eine Warnung an, und Sie werden gefragt, ob die Änderung trotzdem durchgeführt werden soll.

Gehe zu Fall

Mit dem Dialogfeld "Gehe zu Fall" wechseln Sie im Daten-Editor zum Fall (zur Zeile) mit der angegebenen Nummer.

Abbildung 5-7
Dialogfeld "Gehe zu Fall"



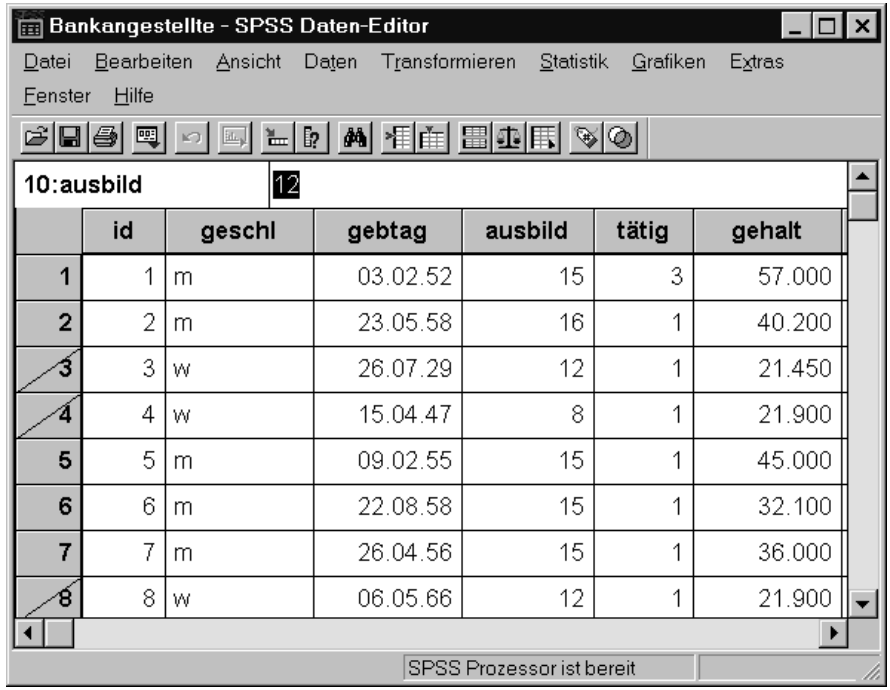
So können Sie einen Fall im Daten-Editor suchen:

- Aktivieren Sie das Fenster des Daten-Editors.
- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Daten
  Gehe zu Fall...
- Geben Sie im Dialogfeld "Gehe zu Fall" die Zeilennummer aus dem Daten-Editor für den Fall ein.

Status für die Fallauswahl im Daten-Editor

Wenn Sie eine Teilmenge von Fällen ausgewählt, nicht ausgewählte Fälle jedoch nicht verworfen haben, sind die nicht ausgewählten Fälle im Daten-Editor mit einer diagonalen Linie (Schrägstrich) durch die Zeilennummer gekennzeichnet.

Abbildung 5-8
Gefilterte Fälle im Daten-Editor



	id	geschl	gebtag	ausbild	tätig	gehalt
1	1	m	03.02.52	15	3	57.000
2	2	m	23.05.58	16	1	40.200
3	3	w	26.07.29	12	1	21.450
4	4	w	15.04.47	8	1	21.900
5	5	m	09.02.55	15	1	45.000
6	6	m	22.08.58	15	1	32.100
7	7	m	26.04.56	15	1	36.000
8	8	w	06.05.66	12	1	21.900

Optionen für die Anzeige im Daten-Editor

Das Menü “Ansicht” bietet verschiedene Anzeigeoptionen für den Daten-Editor:

Schriftarten. Mit dieser Option können Sie die Schrifteigenschaften der Datenanzeige festlegen.

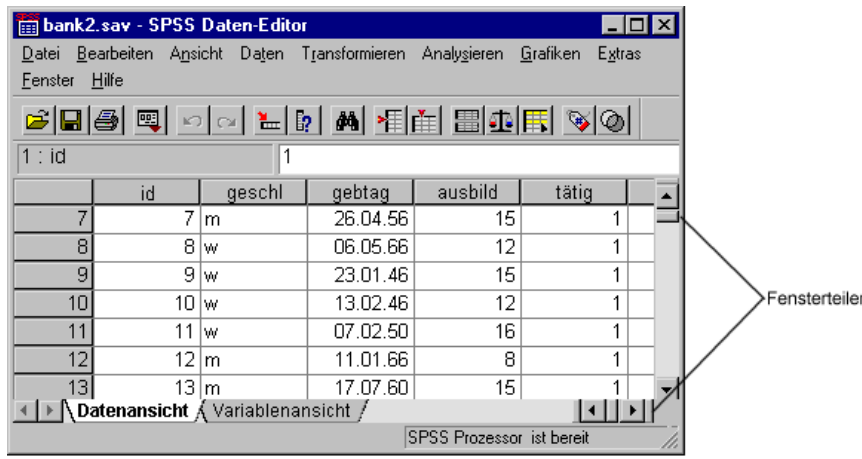
Gitterlinien. Mit dieser Option werden Gitterlinien ein- und ausgeblendet.

Wertelabels. Mit dieser Option wechseln Sie zwischen der Anzeige der tatsächlichen Datenwerte und der benutzerdefinierten beschreibenden Wertelabels hin und her. Diese Option ist nur in der Datenansicht verfügbar.

Verwenden mehrerer Ansichten

In der Datenansicht können Sie mehrere Ansichten (Fensterbereiche) mithilfe der Fensterteiler oberhalb der horizontalen Bildlaufleiste und rechts neben der vertikalen Bildlaufleiste anlegen.

Abbildung 5-9
Fensterteiler in der Datenansicht


bank2.sav - SPSS Daten-Editor
Datei Bearbeiten Ansicht Daten Transformieren Analysieren Grafiken Extras
Fenster Hilfe
1 : id
1
id geschl gebtag ausbild tätig
7 7 m 26.04.56 15 1
8 8 w 06.05.66 12 1
9 9 w 23.01.46 15 1
10 10 w 13.02.46 12 1
11 11 w 07.02.50 16 1
12 12 m 11.01.66 8 1
13 13 m 17.07.60 15 1
Datenansicht
Variablenansicht
SPSS Prozessor ist bereit
Fensterteiler

Des weiteren können Sie Fensterteiler über das Menü “Fenster” einfügen und wieder entfernen. So fügen Sie Fensterteiler ein:

▶ Wählen Sie die folgenden Befehle aus den Menüs der Datenansicht aus:

Fenster
  Aufteilen

Fensterteiler werden oberhalb und links von der ausgewählten Zelle eingefügt.

- Wenn Sie die obere linke Zelle auswählen, werden die Fensterteiler so plaziert, dass die aktuelle Ansicht horizontal und vertikal etwa gleich geteilt wird.
- Wählen Sie nicht die oberste Zelle in der ersten Spalte aus, wird ein horizontaler Fensterteiler oberhalb der ausgewählten Zelle eingefügt.
- Wählen Sie nicht die erste Zelle in der obersten Zeile aus, wird ein vertikaler Fensterteiler links neben der ausgewählten Zelle eingefügt.

Drucken aus dem Daten-Editor

Eine Datendatei wird so gedruckt, wie sie auf dem Bildschirm angezeigt wird.

- Die in der gegenwärtig angezeigten Ansicht enthaltenen Informationen werden gedruckt. In der Datenansicht werden die Daten gedruckt. In der Variablenansicht werden die Informationen zu den Variablendefinitionen gedruckt.
- Gitterlinien werden gedruckt, wenn sie gegenwärtig in der ausgewählten Ansicht angezeigt werden.
- Wenn in der Datenansicht gegenwärtig Wertelabels angezeigt werden, werden diese gedruckt. Andernfalls werden die eigentlichen Datenwerte gedruckt.

Verwenden Sie das Menü “Ansicht” im Fenster des Daten-Editors, um Gitterlinien ein- oder auszublenden und zwischen der Anzeige von Datenwerten bzw. Wertelabels umzuschalten.

So drucken Sie den Inhalt des Daten-Editors

▶ Aktivieren Sie das Fenster des Daten-Editors.

▶ Klicken Sie auf die Registerkarte der Ansicht, die gedruckt werden soll.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Datei
  Drucken...

Kapitel
6

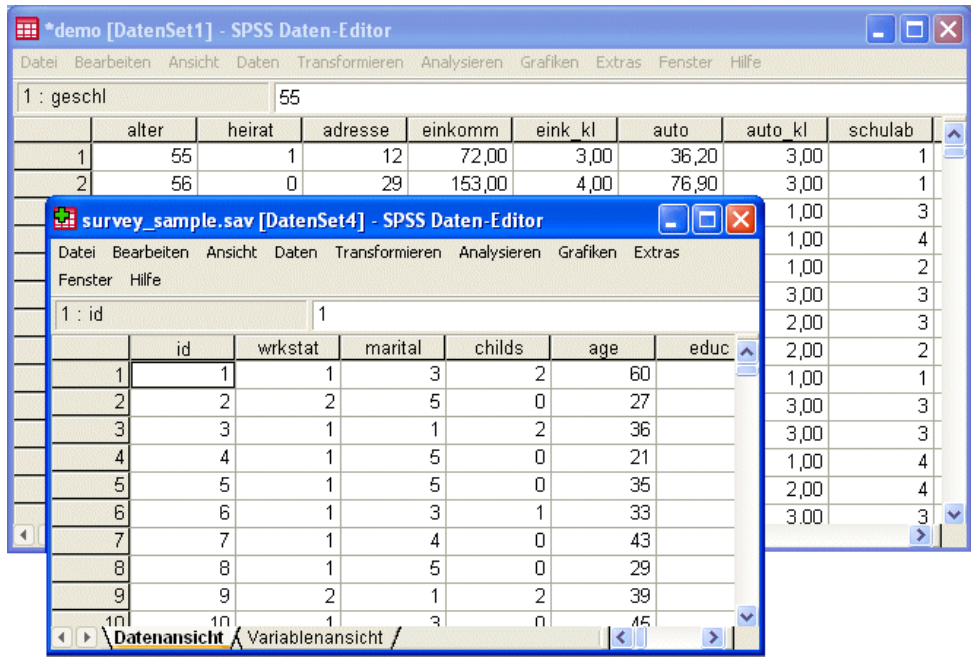
Arbeiten mit mehreren Datenquellen

Ab Version 14.0 können Sie in SPSS mehrere Datenquellen gleichzeitig geöffnet haben. Dies vereinfacht Folgendes:

- Wechseln zwischen Datenquellen
- Vergleichen der verschiedenen Datenquellen
- Kopieren und Einfügen von Daten zwischen Datenquellen
- Erstellen von Teilmengen der Fälle und/oder Variablen für die Analyse
- Zusammenführen von verschiedenen Datenquellen mit unterschiedlichen Datenformaten (beispielsweise Tabellenkalkulationsblätter, Datenbanken, Textdaten), ohne dass zuerst jede Datenquellen im SPSS-Format gespeichert werden muss

Grundsätzlicher Umgang mit mehreren Datenquellen

Abbildung 6-1
Zwei gleichzeitig geöffnete Datenquellen

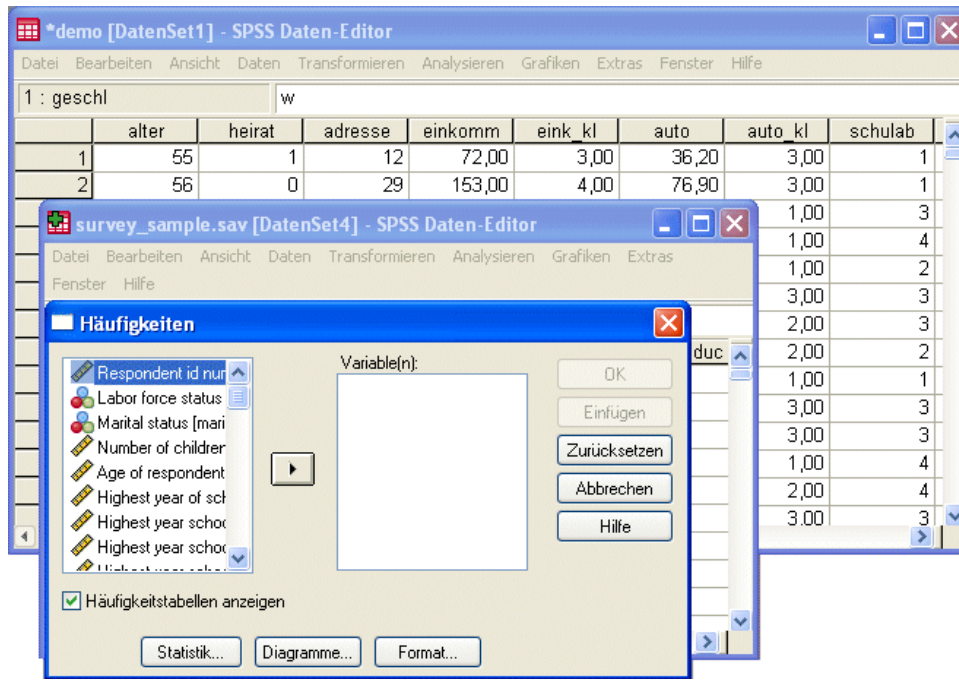

*demo [DatenSet1] - SPSS Daten-Editor
Datei Bearbeiten Ansicht Daten Transformieren Analysieren Grafiken Extras Fenster Hilfe
1 : geschl
55
alter heirat adresse einkomm eink_kl auto auto_kl schulab
1 55 1 12 72,00 3,00 36,20 3,00 1
2 56 0 29 153,00 4,00 76,90 3,00 1
1,00 3
1,00 4
1,00 2
3,00 3
2,00 3
2,00 2
1,00 1
3,00 3
3,00 3
1,00 4
2,00 4
3,00 3
survey_sample.sav [DatenSet4] - SPSS Daten-Editor
Datei Bearbeiten Ansicht Daten Transformieren Analysieren Grafiken Extras
Fenster Hilfe
1 : id
1
id wrkstat marital childs age educ
1 1 1 3 2 60
2 2 2 5 0 27
3 3 1 1 2 36
4 4 1 5 0 21
5 5 1 5 0 35
6 6 1 3 1 33
7 7 1 4 0 43
8 8 1 5 0 29
9 9 2 1 2 39
Datenansicht Variablenansicht

Jede Datenquelle, die Sie öffnen, wird in einem neuen Fenster des Daten-Editors angezeigt.

- Jede zuvor geöffnete Datenquelle bleibt geöffnet und für die weitere Verwendung verfügbar.
- Wenn Sie eine Datenquelle öffnen, wird diese automatisch zur **Arbeitsdatei**.
- Sie können die Arbeitsdatei wechseln, indem Sie auf eine beliebige Stelle im Fenster “Daten-Editor” der gewünschten Datenquelle klicken oder das Fenster “Daten-Editor” für diese Datenquelle aus dem Menü “Fenster” auswählen.

- Nur die Variablen in der Arbeitsdatei sind für die Analyse verfügbar.

Abbildung 6-2
Variablenliste mit Variablen aus der Arbeitsdatei



- Sie können die Arbeitsdatei nicht wechseln, wenn ein Dialogfeld geöffnet ist, mit dem auf die Daten zugegriffen wird (einschließlich aller Dialogfelder, in denen Variablenlisten angezeigt werden).
- Während einer Sitzung muss mindestens ein Fenster des Daten-Editors geöffnet sein. Wenn Sie das letzte geöffnete Fenster des Daten-Editors schließen, wird SPSS automatisch beendet. Dabei werden Sie aufgefordert, die Änderungen zu speichern.

Anmerkung: Wenn Sie die Befehlssyntax verwenden, um Datenquellen zu öffnen (z. B. `GET FILE` oder `GET DATA`), muss jedes Daten-Set eindeutig benannt werden, damit mehr als eine Datenquelle gleichzeitig geöffnet werden kann.

Kopieren und Einfügen von Informationen zwischen Daten-Sets

Sowohl Daten als auch die Attribute zur Variablendefinition werden grundsätzlich auf die gleiche Art aus einem Daten-Set in ein anderes Daten-Set kopieren, in der Sie Informationen innerhalb einer einzelnen Datendatei kopieren und einfügen.

- Beim Kopieren und Einfügen von ausgewählten Datenzellen in der Datenansicht werden nur die Datenwerte eingefügt, nicht die Attribute zur Variablendefinition.
- Beim Kopieren und Einfügen einer vollständigen Variablen in der Datenansicht durch Auswählen der Variablen im Spaltenkopf werden alle Daten und alle Attribute zur Variablendefinition für diese Variable eingefügt.
- Beim Kopieren und Einfügen von Attributen zur Variablendefinition oder vollständigen Variablen in der Variablenansicht werden die ausgewählten Attribute (oder die vollständige Variablendefinition) eingefügt, nicht jedoch die Datenwerte.

Umbenennen von Daten-Sets

Wenn Sie eine Datenquelle mit den Menüs und Dialogfeldern öffnen, wird jeder Datenquelle automatisch der Name *Daten-Setn* zugewiesen, wobei *n* eine fortlaufende Ganzzahl ist. Wenn Sie eine Datenquelle mit der Befehlssyntax öffnen, wird dem Daten-Set kein Name zugewiesen, sofern Sie nicht ausdrücklich mit `DATASET NAME` einen Namen angeben. So vergeben Sie aussagekräftigere Namen für die Daten-Sets:

▶ Wählen Sie folgenden Optionen für das Daten-Set, dessen Namen Sie ändern möchten, aus den Menüs im Fenster des Daten-Editors aus:

Datei
  Daten-Set umbenennen

▶ Geben Sie einen neuen Namen für das Daten-Set ein, der den SPSS-Regeln für Variablennamen entspricht. Für weitere Informationen siehe “Variablennamen” in Kapitel 5 auf S. 99.

Kapitel
7

Aufbereitung von Daten

Wenn Sie Daten im Daten-Editor eingegeben oder eine Datendatei geöffnet haben, können Sie ohne weitere Vorarbeit mit dem Erstellen von Berichten, Diagrammen und Analysen beginnen. Es gibt jedoch noch einige zusätzliche nützliche Funktionen zur Datenvorbereitung. Dazu gehören folgende Möglichkeiten:

- Zuweisen von Variableneigenschaften, welche die Daten beschreiben und festlegen, wie bestimmte Werte behandelt werden sollen.
- Identifizieren von Fällen die eventuell doppelte Informationen enthalten und Auschluß dieser Fälle aus den Analysen oder Löschen der Fälle aus der Datendatei.
- Erstellen neuer Variablen mit einigen verschiedenen Kategorien, die für Wertebereiche aus Variablen mit einer großen Anzahl möglicher Werte stehen.

Variableneigenschaften

Die im Daten-Editor in der Datenansicht eingegebenen Rohdaten oder die aus einem externen Dateiformat in SPSS importierten Daten (zum Beispiel aus einer Excel-Tabellenkalkulationsdatei einer Textdatei) verfügen noch nicht über einige spezielle, möglicherweise sehr nützliche Variableneigenschaften, wie zum Beispiel:

- Definition aussagekräftiger Wertelabels für numerische Codes (beispielsweise 0 = *Männlich* und 1 = *Weiblich*).
- Kennzeichnung fehlender Werte mit Codes (beispielsweise 99 = *Nicht zutreffend*).
- Zuweisung von Meßniveaus (nominal, ordinal oder metrisch).

Diese und viele weitere Variableneigenschaften können in der Variablenansicht des Daten-Editors zugewiesen werden. Darüber hinaus stehen Ihnen zahlreiche Hilfsmittel zur Durchführung dieses Vorgangs zur Verfügung:

- **Variableneigenschaften definieren.** Diese Funktion unterstützt Sie bei der Definition von aussagekräftigen Wertelabels und fehlenden Werten. Sie ist besonders hilfreich bei der Aufbereitung kategorialer Daten mit numerischen Codes für Kategorienwerte.
- **Dateneigenschaften kopieren.** Mit dieser Funktion können Sie eine vorhandene Datendatei im SPSS-Format als Vorlage für die Datei- und Variableneigenschaften in der aktuellen Datendatei verwenden. Dies ist besonders nützlich, wenn Sie regelmäßig externe Datendateien mit ähnlichem Datenbestand verwenden (beispielsweise Monatsberichte im Excel-Format).

Definieren von Variableneigenschaften

Die Funktion "Variableneigenschaften definieren" ist daraufhin konzipiert, Ihnen das Erstellen von beschreibenden Wertelabels für kategoriale, nominale oder ordinale Variablen zu erleichtern. Mit Hilfe dieser Funktion können folgende Vorgänge durchgeführt werden:

- Durchsuchen der tatsächlichen Datenwerte und Auflisten aller eindeutigen Datenwerte für die ausgewählten Variablen.
- Ermitteln der Werte ohne Label und Bereitstellen einer Funktion zur automatischen Beschriftung.
- Kopieren definierter Wertelabels aus einer anderen Variablen in die ausgewählte Variable oder von der ausgewählten Variablen in mehrere zusätzliche Variablen.

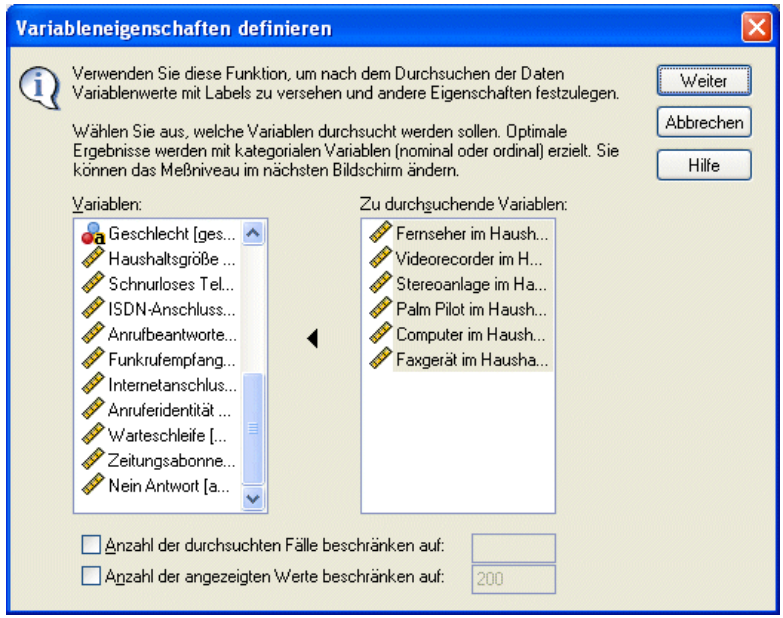
Anmerkung: Wenn von der Funktion "Variableneigenschaften definieren" keine Fälle durchsucht werden sollen, geben Sie "0" als Anzahl der zu durchsuchenden Fälle ein.

So definieren Sie Variableneigenschaften

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Daten
  Variableneigenschaften definieren...

Abbildung 7-1
Erstes Dialogfeld für die Auswahl zu definierender Variablen



- Wählen Sie die numerischen oder kurzen String-Variablen aus, für die Wertelabels erstellt oder andere Variableneigenschaften wie fehlende Werte oder beschreibende Variablenlabels definiert bzw. geändert werden sollen.

  Anmerkung: Lange String-Variablen (String-Variablen mit einer definierten Länge von mehr als acht Zeichen) werden in der Variablenliste nicht angezeigt. Für lange String-Variablen können keine Wertelabels oder Kategorien für fehlende Werte definiert werden.

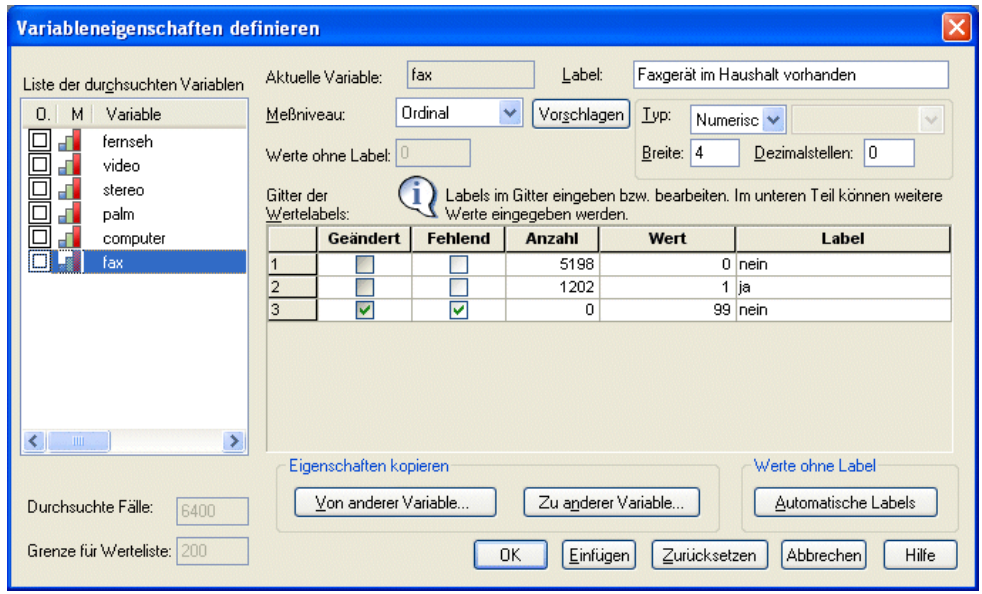
- Geben Sie die Anzahl der Fälle an, die durchsucht und bei Erstellung einer Liste mit eindeutigen Werten berücksichtigt werden sollen. Dies ist insbesondere bei Datendateien mit einer großen Anzahl an Fällen nützlich, da sich das Durchsuchen der gesamten Datendatei in diesem Fall sehr zeitaufwendig gestalten würde.

- Geben Sie eine Obergrenze für die Anzahl der eindeutigen Werte an, die angezeigt werden soll. Durch diese Angabe wird in erster Linie vermieden, daß Hunderte, Tausende oder sogar Millionen von Werten für metrische Variablen bzw. stetige Intervall- oder Verhältnisvariablen aufgelistet werden.

- Klicken Sie auf Weiter, um das Hauptdialogfeld “Variableneigenschaften definieren” zu öffnen.
- Wählen Sie eine Variable aus, für die Wertelabels erstellt oder andere Variableneigenschaften definiert bzw. geändert werden sollen.
- Geben Sie den Beschriftungstext für alle Werte ohne Label ein, die im “Gitter der Wertelabels” angezeigt werden.
- Wenn Wertelabels für Werte erstellt werden sollen, die nicht angezeigt werden, können Sie die Werte in der Spalte *Werte* unter dem letzten durchsuchten Wert eingeben.
- Wiederholen Sie diesen Vorgang für jede aufgeführte Variable, für die Wertelabels erstellt werden sollen.
- Klicken Sie auf OK, um die Wertelabels und Variableneigenschaften zuzuweisen.

Definieren von Wertelabels und anderen Variableneigenschaften

Abbildung 7-2
Hauptdialogfeld “Variableneigenschaften definieren”



Das Hauptdialogfeld "Variableneigenschaften definieren" enthält folgende Informationen über die durchsuchten Variablen:

Liste der durchsuchten Variablen. Wenn eine durchsuchte Variable Werte ohne zugewiesene Wertelabels enthält, wird dies durch ein Häkchen in der Spalte *Ohne Label* angezeigt.

So sortieren Sie die Variablenliste, um alle Variablen mit Werten ohne Label am Anfang der Liste anzuzeigen:

▶ Klicken Sie in der Liste der durchsuchten Variablen auf die Spaltenüberschrift *Ohne Label*.

Sie können auch nach Variablennamen oder Meßniveau sortieren, indem Sie in der Liste der durchsuchten Variablen auf die entsprechende Spaltenüberschrift klicken.

Gitter der Wertelabels

- **Label (Beschriftung).** Zeigt alle bereits definierten Wertelabels an. In dieser Spalte können Sie Labels hinzufügen oder ändern.
- **Wert.** Zeigt für jede ausgewählte Variable die eindeutigen Werte an. Diese Liste mit eindeutigen Werten beruht auf der Anzahl der durchsuchten Fälle. Wenn beispielsweise nur die ersten 100 Fälle in der Datendatei durchsucht wurden, gibt die Liste nur die in diesen Fällen auftretenden eindeutigen Werte wieder. Wenn die Datendatei zuvor nach der Variablen sortiert wurde, der Wertelabels zugewiesen werden sollen, werden in der Liste möglicherweise weitaus weniger eindeutige Werte angezeigt, als tatsächlich in den Daten vorhanden sind.
- **Häufigkeiten.** Die Häufigkeit, mit der jeder Wert in den durchsuchten Fällen auftritt.
- **Fehlend.** Werte, für die definiert wurde, daß sie fehlende Daten darstellen. Sie können die Zuweisung fehlender Werte für die Kategorie ändern, indem Sie auf das Kontrollkästchen klicken. Ein Häkchen zeigt an, daß die Kategorie als benutzerdefinierte fehlende Kategorie definiert ist. Wenn für eine Variable bereits ein Bereich von Werten als benutzerdefiniert fehlend definiert ist (z. B. 90-99), können Sie für diese Variable mit Hilfe der Funktion "Variableneigenschaften definieren" keine Kategorien für fehlende Werte hinzufügen oder ändern. Für Variablen mit Bereichen von fehlenden Werten können die Kategorien für

fehlende Werte in der Variablenansicht des Daten-Editors geändert werden. Für weitere Informationen siehe “Fehlende Werte” in Kapitel 5 auf S. 106.

- **Geändert.** Zeigt an, daß ein Wertelabel hinzugefügt oder geändert wurde.

Anmerkung: Wenn Sie im ersten Dialogfeld als Anzahl der zu durchsuchenden Fälle “0” angegeben haben, ist das Gitter der Wertelabels mit Ausnahme der für die Variable bereits definierten Wertelabels und/oder Kategorien für fehlende Werte zu Beginn leer. Außerdem ist die Schaltfläche Vorschlagen für das Meßniveau deaktiviert.

Meßniveau. Wertelabels sind in erster Linie für kategoriale, d. h. nominale und ordinale Variablen, sinnvoll. Zudem werden kategoriale Variablen in einigen Prozeduren anders behandelt als metrische Variablen, sodass das Zuweisen des richtigen Meßniveaus unter Umständen wichtig ist. In der Standardeinstellung wird allen neuen numerischen Variablen das metrische Meßniveau zugewiesen. Daher werden möglicherweise auch viele kategoriale Variablen zunächst als metrisch angezeigt.

Wenn Sie sich nicht sicher sind, welches Meßniveau einer Variablen zugewiesen werden soll, klicken Sie auf Vorschlagen.

Eigenschaften kopieren. Sie können Wertelabels und andere Variableneigenschaften aus einer anderen Variablen in die gerade ausgewählte Variable oder aus der gerade ausgewählten Variablen in eine oder mehr andere Variablen kopieren.

Werte ohne Label. Zum automatischen Erstellen von Beschriftungen für Werte ohne Labels klicken Sie auf Automatische Labels.

Variablenlabel und Anzeigeformat

Sie können das beschreibende Variablenlabel und das Anzeigeformat ändern.

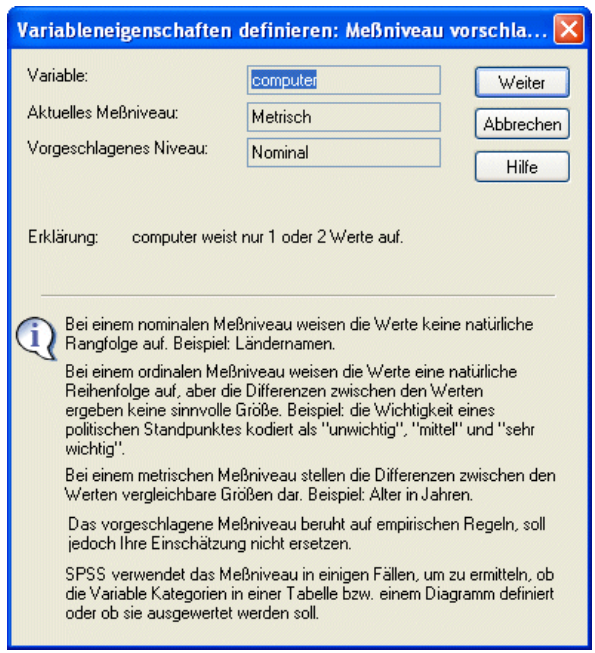
- Der grundlegende Typ der Variablen kann jedoch nicht geändert werden (String oder numerisch).
- Bei String-Variablen können Sie nur das Variablenlabel ändern, nicht jedoch das Anzeigeformat.
- Bei numerischen Variablen können Sie den numerischen Typ (z. B. numerisch, Datum, Dollar oder spezielle Währung), die Breite (Höchstzahl der Ziffern einschließlich aller Dezimal- oder Gruppentrennzeichen) sowie die Anzahl der Dezimalstellen ändern.

- Beim numerischen Datumsformat können Sie ein bestimmtes Datumsformat auswählen (z. B. tt-mm-jjjj, mm/tt/jj, jjjjttt usw.).
- Bei benutzerdefinierten numerischen Formaten können Sie eine von fünf benutzerdefinierten Währungsformaten auswählen (CCA bis CCE). Für weitere Informationen siehe “Optionen: Währung” in Kapitel 44 auf S. 683.
- In der Spalte *Wert* wird ein Sternchen (*) angezeigt, wenn die angegebene Breite kleiner ist als die Breite der durchsuchten Werte oder der Werte für bereits vorhandene definierte Wertelabels bzw. für Kategorien für fehlende Werte.
- Ein Punkt (.) wird angezeigt, wenn die durchsuchten Werte oder die Werte für bereits vorhandene definierte Wertelabels bzw. Kategorien für fehlende Werte für den ausgewählten Anzeigetyp ungültig sind. Ein interner numerischer Wert von weniger als 86.400 ist beispielsweise für eine Variable im Datumsformat ungültig.

Zuweisen des Meßniveaus

Wenn Sie im Hauptdialogfeld “Variableneigenschaften definieren” für das Meßniveau auf Vorschlagen klicken, wird die aktuelle Variable anhand der durchsuchten Fälle und der definierten Wertelabels bewertet. Anschließend wird das Dialogfeld “Meßniveau vorschlagen” mit einem Vorschlag für das Meßniveau eingeblendet. Im Bereich “Erklärung” finden Sie eine kurze Beschreibung der Kriterien, nach denen das vorgeschlagene Meßniveau ausgewählt wurde.

Abbildung 7-3
Dialogfeld "Meßniveau vorschlagen"


Variableneigenschaften definieren: Meßniveau vorschla...
Variable:
computer
Aktuelles Meßniveau:
Metrisch
Vorgeschlagenes Niveau:
Nominal
Weiter
Abbrechen
Hilfe
Erklärung:
computer weist nur 1 oder 2 Werte auf.
Bei einem nominalen Meßniveau weisen die Werte keine natürliche Rangfolge auf. Beispiel: Ländernamen.
Bei einem ordinalen Meßniveau weisen die Werte eine natürliche Reihenfolge auf, aber die Differenzen zwischen den Werten ergeben keine sinnvolle Größe. Beispiel: die Wichtigkeit eines politischen Standpunktes kodiert als "unwichtig", "mittel" und "sehr wichtig".
Bei einem metrischen Meßniveau stellen die Differenzen zwischen den Werten vergleichbare Größen dar. Beispiel: Alter in Jahren.
Das vorgeschlagene Meßniveau beruht auf empirischen Regeln, soll jedoch Ihre Einschätzung nicht ersetzen.
SPSS verwendet das Meßniveau in einigen Fällen, um zu ermitteln, ob die Variable Kategorien in einer Tabelle bzw. einem Diagramm definiert oder ob sie ausgewertet werden soll.

Anmerkung: Werte, die als fehlende Werte definiert wurden, werden bei der Berechnung des Meßniveaus nicht berücksichtigt. In der Erklärung für das vorgeschlagene Meßniveau kann beispielsweise darauf hingewiesen werden, daß der Vorschlag teilweise darauf beruht, daß die Variable keine negativen Werte enthält, obgleich sie in Wirklichkeit möglicherweise negative Werte enthält, die jedoch bereits als fehlende Werte definiert sind.

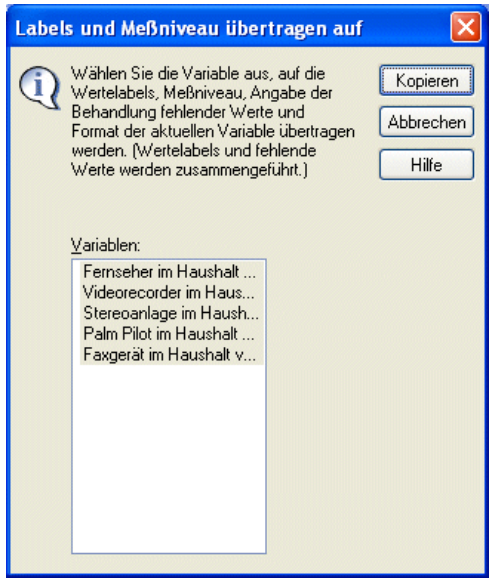
- Klicken Sie auf Weiter, um das vorgeschlagene Meßniveau zu übernehmen, oder auf Abbrechen, um das Meßniveau unverändert beizubehalten.

Kopieren von Variableneigenschaften

Das Dialogfeld "Labels und Meßniveau übertragen" wird angezeigt, wenn Sie im Hauptdialogfeld "Variableneigenschaften definieren" auf Von anderer Variable oder Zu anderer Variable klicken. In diesem Dialogfeld werden alle durchsuchten Variablen

angezeigt, die mit dem Typ der aktuellen Variablen übereinstimmen (numerisch oder String). Bei String-Variablen muß auch die definierte Länge übereinstimmen.

Abbildung 7-4
Dialogfeld "Labels und Meßniveau übertragen"


Labels und Meßniveau übertragen auf
Wählen Sie die Variable aus, auf die Wertelabels, Meßniveau, Angabe der Behandlung fehlender Werte und Format der aktuellen Variable übertragen werden. (Wertelabels und fehlende Werte werden zusammengeführt.)
Kopieren
Abbrechen
Hilfe
Variablen:
Fernseher im Haushalt ...
Videorecorder im Haus...
Stereoanlage im Haush...
Palm Pilot im Haushalt ...
Faxgerät im Haushalt v...

- Wählen Sie eine Variable aus, von der Wertelabels und andere Variableneigenschaften (außer dem Variablenlabel) kopiert werden sollen.

  oder

- Wählen Sie eine oder mehrere Variablen aus, in die Wertelabels und andere Variableneigenschaften kopiert werden sollen.

- Klicken Sie auf Kopieren, um die Wertelabels und das Meßniveau zu kopieren.
  - Wertelabels und Kategorien für fehlende Werte, die bereits in der Zielvariablen vorhanden sind, werden nicht ersetzt.
  - Wertelabels und Kategorien für fehlende Werte, die noch nicht für die Zielvariable(n) definiert sind, werden den Wertelabels und den Kategorien für fehlende Werte der Zielvariablen hinzugefügt.

- Das Meßniveau für die Zielvariable(n) wird immer ersetzt.
- Wenn entweder die Quell- oder die Zielvariable einen definierten Bereich von fehlenden Werten aufweist, werden die Definitionen für die fehlenden Werte nicht kopiert.

Mehrfachantworten-Sets

In benutzerdefinierten Tabellen und der Diagrammerstellung kann eine besondere Art von "Variable" verwendet werden, die als **Mehrfachantworten-Set** bezeichnet wird. Bei Mehrfachantworten-Sets handelt es sich nicht um "Variablen" im üblichen Sinn. Mehrfachantworten-Sets können nicht im Daten-Editor angezeigt werden, sie werden von anderen Prozeduren nicht erkannt. Mehrfachantworten-Sets verwenden mehrere Variablen, um Antworten auf Fragen aufzuzeichnen, auf welche der Befragte mehr als eine Antwort geben kann. Sie werden wie kategoriale Variablen behandelt und bieten weitestgehend dieselben Möglichkeiten wie kategoriale Variablen.

Mehrfachantworten-Sets werden aus mehreren Variablen in der Datendatei gebildet. Beim Mehrfachantworten-Set handelt es sich um ein spezielles Konstrukt innerhalb einer Datendatei im SPSS-Format. Mehrfachantworten-Sets können in einer Datendatei im SPSS-Format definiert und gespeichert werden. Das Importieren oder Exportieren von Mehrfachantworten-Sets in bzw. aus anderen Dateiformaten ist jedoch nicht möglich. (Mit der Option "Dateneigenschaften kopieren" können Sie Mehrfachantworten-Sets aus anderen SPSS-Datendateien kopieren. Sie finden diese Option im Menü "Daten" im Fenster des Daten-Editors.Für weitere Informationen siehe "Kopieren von Dateneigenschaften" auf S. 138.)

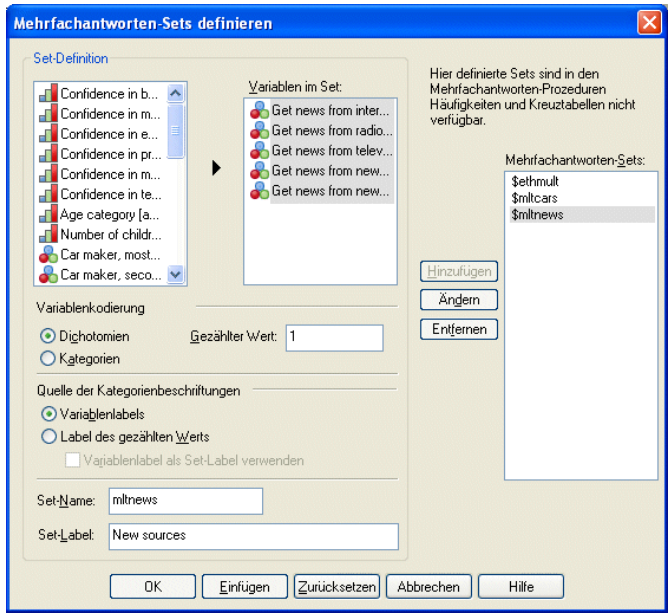
Definieren von Mehrfachantworten-Sets

So definieren Sie Mehrfachantworten-Sets:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Daten
  Mehrfachantworten-Sets definieren...

Abbildung 7-5
Dialogfeld "Mehrfachantworten-Sets"



- Wählen Sie mindestens zwei Variablen aus. Wenn Ihre Variablen als Dichotomien kodiert sind, geben Sie an, welcher Wert gezählt werden soll.
- Geben Sie einen eindeutigen Namen für jedes Mehrfachantworten-Set ein. Der Name darf maximal 63 Byte lang sein. Dem Set-Namen wird automatisch ein Dollarzeichen vorangestellt.
- Geben Sie einen aussagekräftigen Namen für das Set ein. (Optional.)
- Klicken Sie auf Hinzufügen, um das Mehrfachantworten-Set zur Liste der definierten Sets hinzuzufügen.

Dichotomien

Ein Set aus dichotomen Variablen enthält gewöhnlich mehrere dichotome Variablen, d. h. Variablen mit nur zwei möglichen Werten der folgenden Art: ja/nein, vorhanden/nicht vorhanden oder angekreuzt/nicht angekreuzt. Wenngleich die Variablen nicht immer streng dichotom sein müssen, werden alle Variablen des

Sets auf dieselbe Weise kodiert, und für "Gezählter Wert" gilt die Bedingung "positiv/vorhanden/angekreuzt".

In einer Umfrage wird beispielsweise die folgende Frage gestellt: "Aus welchen der folgenden Quellen beziehen Sie Nachrichten?", wobei fünf Antwortmöglichkeiten zur Auswahl stehen. Die Befragten können mehrere Antworten geben, indem sie mehrere Felder neben den einzelnen Antworten ankreuzen. Die fünf Antworten entsprechen fünf Variablen in der Datendatei, wobei 0 für *Nein* (nicht angekreuzt) und 1 für *Ja* (angekreuzt) steht. Im Set aus dichotomen Variablen ist der "Gezählte Wert" gleich 1.

Die Stichprobendatendatei *survey_sample.sav* (im Unterverzeichnis *tutorial/sample_files* des Installationsverzeichnisses) enthält bereits drei definierte Mehrfachantworten-Sets. *$mltnews* ist ein Set aus dichotomen Variablen.

▶ Wählen Sie *$mltnews* in der Liste für Mehrfachantworten-Sets aus, indem Sie darauf klicken.

Daraufhin werden die Variablen und Einstellungen angezeigt, die für die Definition dieses Mehrfachantworten-Sets verwendet wurden.

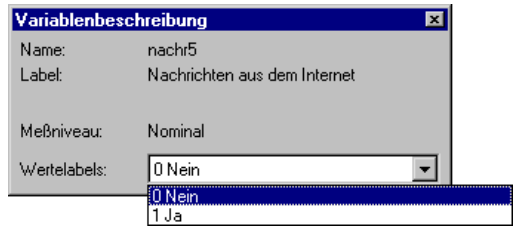
- Die Liste "Variablen im Set" zeigt die fünf Variablen an, mit denen das Mehrfachantworten-Set erstellt wurde.
- Im Gruppenfeld "Variablen kodiert als" wird angezeigt, dass es sich bei den Variablen um dichotome Variablen handelt.
- Für "Gezählter Wert" wird 1 angegeben.

▶ Wählen Sie eine Variable in der Liste "Variablen im Set" aus, indem Sie darauf klicken.

▶ Klicken Sie mit der rechten Maustaste auf die Variable, und wählen Sie im Kontextmenü die Option Variablenbeschreibung aus.

▶ Klicken Sie im Fenster "Variablenbeschreibung" auf den Abwärtspfeil neben der Dropdown-Liste "Wertelabels", um die Liste mit den definierten Wertelabels vollständig anzuzeigen.

Abbildung 7-6
Variablenbeschreibung für ein Set aus dichotomen Quellvariablen



Die Wertelabels geben an, dass es sich bei der Variablen um eine dichotome Variable mit den Werten 0 und 1 handelt, die jeweils für *Nein* und *Ja* stehen. Alle fünf Variablen in der Liste sind auf dieselbe Weise kodiert, und der Wert 1 (Code für *Ja*) bildet den gezählten Wert für das Set aus dichotomen Variablen.

Kategorien

Ein Set aus kategorialen Variablen besteht aus Mehrfachvariablen, die alle auf dieselbe Weise kodiert wurden, häufig mit zahlreichen möglichen Antwortkategorien. Ein Umfragethema lautet beispielsweise “Nennen Sie bis zu drei Nationalitäten, die am besten Ihre ethnische Herkunft beschreiben”. Zu diesem Thema gibt es Hunderte von möglichen Antworten. Für die Kodierung wird die Liste jedoch auf die 40 häufigsten Nationalitäten begrenzt und alle anderen auf die Kategorie “Andere” verwiesen. In der Datendatei werden die Auswahlmöglichkeiten zu drei Variablen, wobei jede über 41 Kategorien verfügt (40 kodierte Nationalitäten und eine Kategorie “Andere”).

In der Beispieldatendatei bilden *$ethmult* und *$mltcars* Sets aus kategorialen Variablen.

Quelle der Kategorienbeschriftungen

Bei mehreren Dichotomien können Sie festlegen, wie die Sets beschriftet werden.

- **Variablenlabels.** Hierbei werden die definierten Variablenlabels (oder Variablennamen für Variablen ohne definierte Variablenlabels) als Beschriftungen für die Kategorien des Sets verwendet. Wenn zum Beispiel alle Variablen im Set dasselbe Wertelabel (oder keine Wertelabels) für den gezählten Wert aufweisen (z. B. *Ja*), sollten Sie die Variablenlabels als Beschriftungen für die Kategorien des Sets verwenden.

- **Label des gezählten Werts.** Hierbei werden die definierten Wertelabels der gezählten Werte als Beschriftungen für die Kategorien des Sets verwendet. Wählen Sie diese Option nur aus, wenn alle Variablen ein definiertes Wertelabel für den gezählten Wert aufweisen und sich das Wertelabel für den gezählten Wert in jeder Variable unterscheidet.
- **Variablenlabel als Set-Label verwenden.** Wenn Sie Label des gezählten Werts auswählen, können Sie auch das Variablenlabel der ersten Variablen im Set mit einem definierten Variablenlabel als Set-Label verwenden. Wenn keine der Variablen im Set definierte ein definiertes Variablenlabel aufweist, wird der Name der ersten Variable im Set als Set-Label verwendet.

Kopieren von Dateneigenschaften

Mit dem Assistenten zum Kopieren von Dateneigenschaften können Sie eine externe SPSS-Datendatei als Vorlage für die Definition von Datei- und Variableneigenschaften in der Arbeitsdatei verwenden. Sie können außerdem Variablen in der Arbeitsdatei als Vorlagen für andere Variablen in der Arbeitsdatei verwenden. Sie verfügen über folgende Möglichkeiten:

- Kopieren ausgewählter Dateieigenschaften aus einer externen Datendatei oder einem geöffneten Daten-Set in die Arbeitsdatei. Zu den Dateieigenschaften gehören Dokumente, Dateilabel, Mehrfachantworten-Sets, Variablen-Sets und Gewichtung.
- Kopieren ausgewählter Variableneigenschaften aus einer externen Datendatei oder einem geöffneten Daten-Set in entsprechende Variablen in der Arbeitsdatei. Zu den Variableneigenschaften gehören Wertelabels, fehlende Werte, Meßniveau, Variablenlabels, Druck- und Schreibformate, Ausrichtung und Spaltenbreite (im Daten-Editor).
- Kopieren von ausgewählten Variableneigenschaften einer Variablen in einer externen Datendatei, einem geöffneten Daten-Set oder der Arbeitsdatei in viele Variablen in der Arbeitsdatei.
- Erstellen neuer Variablen in der Arbeitsdatei anhand von ausgewählten Variablen in einer externen Datendatei oder einem geöffneten Daten-Set.

Beim Kopieren von Dateneigenschaften gelten die folgenden allgemeinen Regeln:

- Wenn eine externe Datendatei als Quelldatendatei verwendet werden soll, muß diese Datendatei SPSS-Format aufweisen.

- Wenn die Arbeitsdatei als Quelldatendatei verwendet wird, muss die Datei mindestens eine Variable enthalten. Vollständig leere Arbeitsdateien können nicht als Quelldatendateien verwendet werden.
- Bereits definierte Eigenschaften in der Arbeitsdatei können nicht von undefinierten (leeren) Eigenschaften aus dem Quelldaten-Set überschrieben werden.
- Variableneigenschaften aus Quellvariablen können nur in Zielvariablen mit folgendem Datentyp kopiert werden: String (alphanumerisch) oder numerisch (einschließlich numerisch, Datum und Währung).

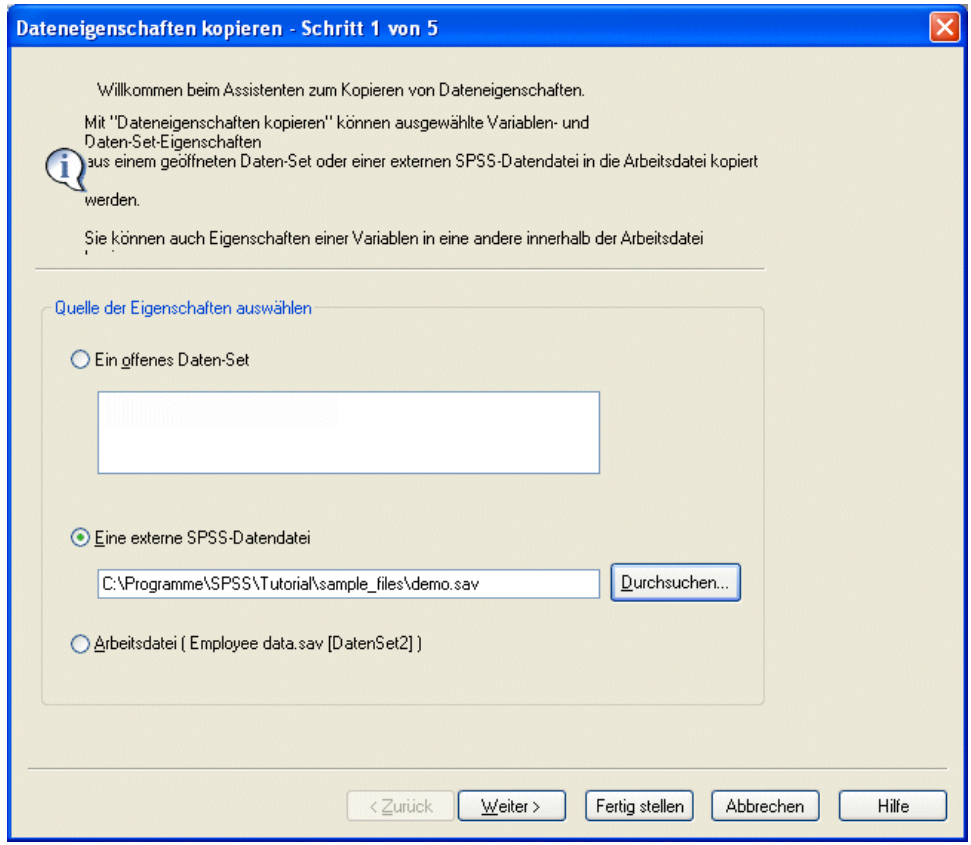
Anmerkung: Die Option “Dateneigenschaften kopieren” ersetzt die Option “Datenlexikon zuweisen”, die in der Vorversion im Menü “Datei” verfügbar war.

So kopieren Sie Dateneigenschaften:

▶ Wählen Sie die folgenden Befehle aus den Menüs im Fenster “Daten-Editor” aus:

Daten
  Dateneigenschaften kopieren...

Abbildung 7-7
Assistent zum Kopieren von Dateneigenschaften: Schritt 1

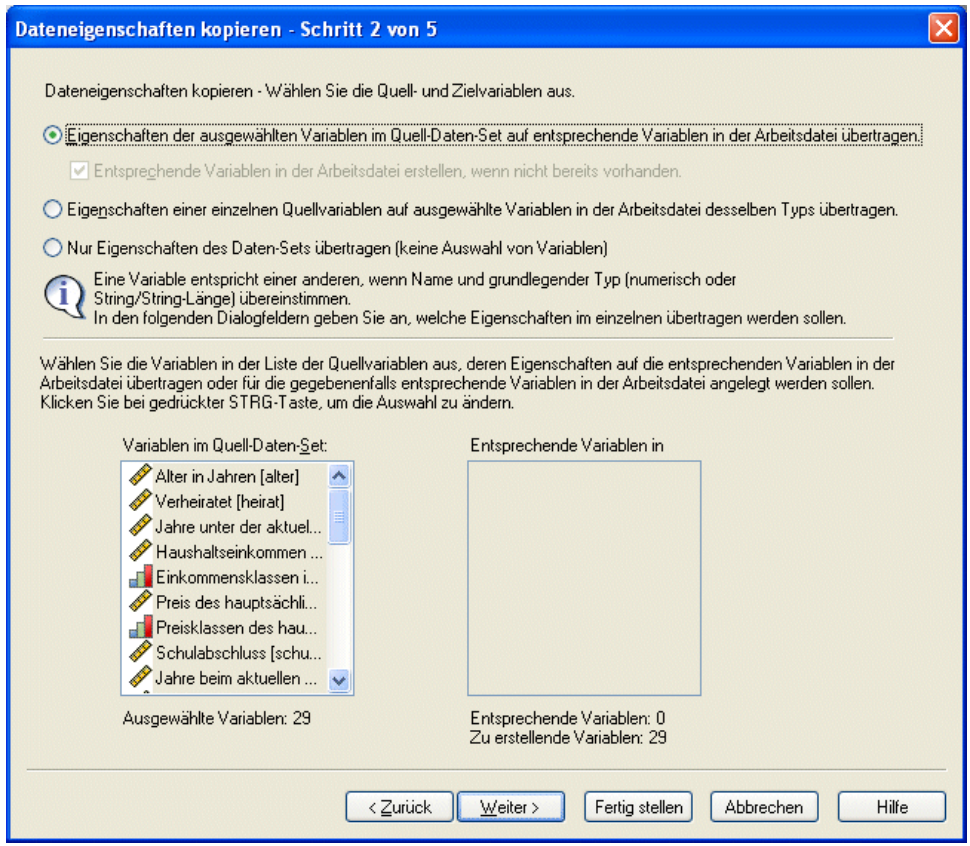

Dateneigenschaften kopieren - Schritt 1 von 5
Willkommen beim Assistenten zum Kopieren von Dateneigenschaften.
Mit "Dateneigenschaften kopieren" können ausgewählte Variablen- und Daten-Set-Eigenschaften
aus einem geöffneten Daten-Set oder einer externen SPSS-Datendatei in die Arbeitsdatei kopiert werden.
Sie können auch Eigenschaften einer Variablen in eine andere innerhalb der Arbeitsdatei
Quelle der Eigenschaften auswählen
Ein offenes Daten-Set
Eine externe SPSS-Datendatei
C:\Programme\SPSS\Tutorial\sample_files\demo.sav
Durchsuchen...
Arbeitsdatei (Employee data.sav [DatenSet2])
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

- Wählen Sie die Datendatei mit den zu kopierenden Datei- und/oder Variableneigenschaften aus. Dabei kann es sich um ein gegenwärtig geöffnetes Daten-Set, eine externe Datendatei im SPSS-Format oder die Arbeitsdatei handeln.
- Befolgen Sie die Anweisungen im Assistenten zum Kopieren von Dateneigenschaften Schritt für Schritt.

Auswählen von Quell- und Zielvariablen

In diesem Schritt können Sie die Quellvariablen mit den zu kopierenden Variableneigenschaften und die Zielvariablen angeben, die diese Variableneigenschaften erhalten sollen.

Abbildung 7-8
Assistent zum Kopieren von Dateneigenschaften: Schritt 2


Dateneigenschaften kopieren - Schritt 2 von 5
Dateneigenschaften kopieren - Wählen Sie die Quell- und Zielvariablen aus.
Eigenschaften der ausgewählten Variablen im Quell-Daten-Set auf entsprechende Variablen in der Arbeitsdatei übertragen.
Entsprechende Variablen in der Arbeitsdatei erstellen, wenn nicht bereits vorhanden.
Eigenschaften einer einzelnen Quellvariablen auf ausgewählte Variablen in der Arbeitsdatei desselben Typs übertragen.
Nur Eigenschaften des Daten-Sets übertragen (keine Auswahl von Variablen)
Eine Variable entspricht einer anderen, wenn Name und grundlegender Typ (numerisch oder String/String-Länge) übereinstimmen.
In den folgenden Dialogfeldern geben Sie an, welche Eigenschaften im einzelnen übertragen werden sollen.
Wählen Sie die Variablen in der Liste der Quellvariablen aus, deren Eigenschaften auf die entsprechenden Variablen in der Arbeitsdatei übertragen oder für die gegebenenfalls entsprechende Variablen in der Arbeitsdatei angelegt werden sollen.
Klicken Sie bei gedrückter STRG-Taste, um die Auswahl zu ändern.
Variablen im Quell-Daten-Set:
Alter in Jahren [alter]
Verheiratet [heirat]
Jahre unter der aktuel...
Haushaltseinkommen ...
Einkommensklassen i...
Preis des hauptsächli...
Preisklassen des hau...
Schulabschluss [schu...
Jahre beim aktuellen ...
Entsprechende Variablen in
Ausgewählte Variablen: 29
Entsprechende Variablen: 0
Zu erstellende Variablen: 29
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

Eigenschaften der ausgewählten Variablen in der Quelldatei auf entsprechende Variablen in der Arbeitsdatei übertragen. Die Variableneigenschaften werden aus einer oder mehreren ausgewählten Quellvariablen in entsprechende Variablen in der Arbeitsdatei kopiert. Variablen "stimmen überein", wenn sowohl der Variablenname als auch der Variablentyp (String oder numerisch) identisch sind. Bei String-Variablen muß

auch die definierte Länge übereinstimmen. In der Standardeinstellung werden in den beiden Variablenlisten nur übereinstimmende Variablen angezeigt.

- **Entsprechende Variablen in der Arbeitsdatei erstellen, wenn nicht bereits vorhanden.** Dadurch wird die Liste mit Quellvariablen so aktualisiert, daß alle Variablen der Quelldatendatei angezeigt werden. Wenn Sie Quellvariablen auswählen, die (dem Variablennamen nach) nicht in der Arbeitsdatei vorhanden sind, werden in der Arbeitsdatei neue Variablen mit den Variablennamen und -eigenschaften aus der Quelldatendatei erstellt.

Wenn die Arbeitsdatei keine Variablen enthält (ein leeres, neues Daten-Set), werden alle Variablen in der Quelldatendatei angezeigt, und in der Arbeitsdatei werden automatisch neue Variablen anhand der ausgewählten Quellvariablen erstellt.

Eigenschaften einer einzelnen Quellvariablen auf ausgewählte Variablen in der Arbeitsdatei desselben Typs übertragen. Die Variableneigenschaften einer in der Liste mit Quellvariablen ausgewählten Variablen können auf mehrere in der Liste "Variablen in Arbeitsdatei" ausgewählte Variablen übertragen werden. In der Liste "Variablen in Arbeitsdatei" werden nur Variablen angezeigt, deren Typ mit der in der Liste "Variablen in Quelldatei" ausgewählten Variablen übereinstimmt (numerisch oder String). Bei einer String-Variablen werden nur Strings angezeigt, deren definierte Länge mit der aus der Quellvariablen übereinstimmt. Diese Option ist nicht verfügbar, wenn die Arbeitsdatei keine Variablen enthält.

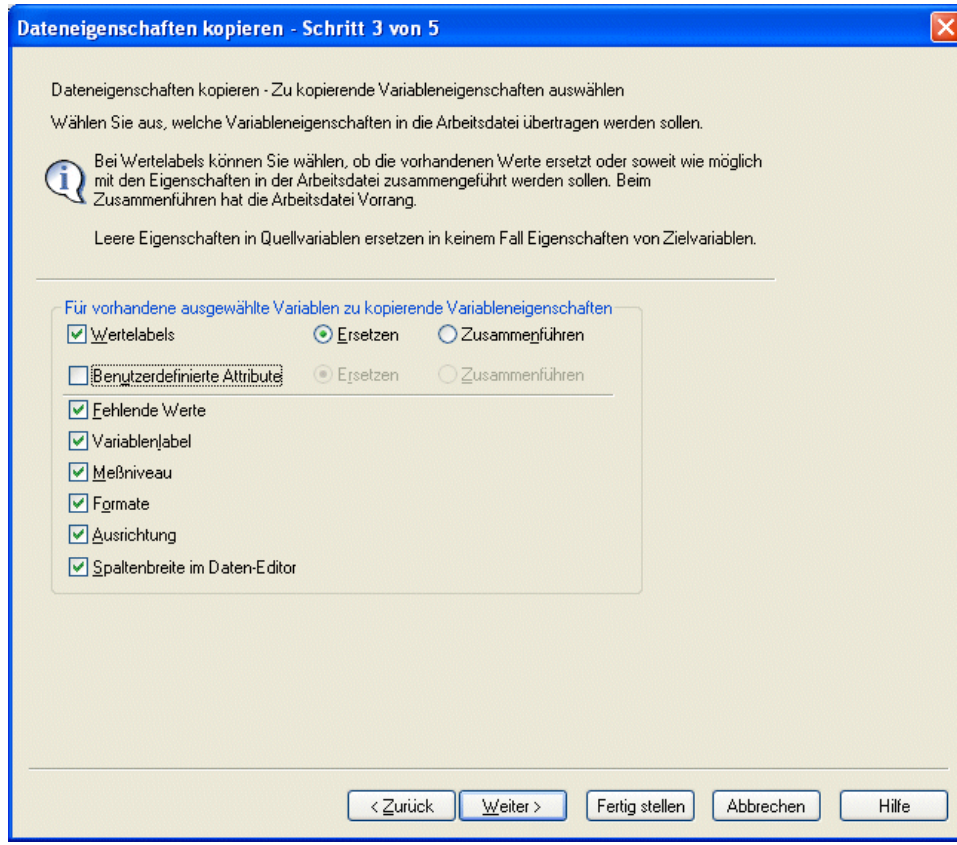
Anmerkung: Mit dieser Option können Sie keine neuen Variablen in der Arbeitsdatei erstellen.

Nur Eigenschaften des Daten-Sets übertragen, keine Auswahl von Variablen. In die Arbeitsdatei werden nur die Dateieigenschaften übertragen (z. B. Dokumente, Dateilabel, Gewichtung). Es werden keine Variableneigenschaften übertragen. Diese Option ist nicht verfügbar, wenn die Arbeitsdatei gleichzeitig die Quelldatendatei darstellt.

Auswählen von Variableneigenschaften zum Kopieren

Sie können ausgewählte Variableneigenschaften aus den Quellvariablen in die Zielvariablen kopieren. Bereits definierte Eigenschaften in den Zielvariablen können nicht von undefinierten (leeren) Eigenschaften aus den Quellvariablen überschrieben werden.

Abbildung 7-9
Assistent zum Kopieren von Dateneigenschaften: Schritt 3


Dateneigenschaften kopieren - Schritt 3 von 5
Dateneigenschaften kopieren - Zu kopierende Variableneigenschaften auswählen
Wählen Sie aus, welche Variableneigenschaften in die Arbeitsdatei übertragen werden sollen.
Bei Wertelabels können Sie wählen, ob die vorhandenen Werte ersetzt oder soweit wie möglich mit den Eigenschaften in der Arbeitsdatei zusammengeführt werden sollen. Beim Zusammenführen hat die Arbeitsdatei Vorrang.
Leere Eigenschaften in Quellvariablen ersetzen in keinem Fall Eigenschaften von Zielvariablen.
Für vorhandene ausgewählte Variablen zu kopierende Variableneigenschaften
Wertelabels
Ersetzen
Zusammenführen
Benutzerdefinierte Attribute
Ersetzen
Zusammenführen
Fehlende Werte
Variablenlabel
Meßniveau
Formate
Ausrichtung
Spaltenbreite im Daten-Editor
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

Wertelabels. Wertelabels sind umschreibende Beschriftungen, die Datenwerten zugeordnet sind. Wertelabels werden häufig verwendet, wenn numerische Datenwerte zur Darstellung nichtnumerischer Kategorien verwendet werden (z. B. die Codes 1 und 2 für *Männlich* und *Weiblich*). Sie können Wertelabels in den Zielvariablen ersetzen oder zusammenführen.

- Durch Ersetzen werden alle definierten Wertelabels der Zielvariablen gelöscht und durch die definierten Wertelabels aus der Quellvariablen ersetzt.
- Durch Zusammenführen werden die definierten Wertelabels der Quellvariablen mit allen definierten Wertelabels der Zielvariablen zusammengeführt. Wenn für denselben Wert sowohl in den Quell- als auch in den Zielvariablen ein Wertelabel definiert ist, wird das Wertelabel in der Zielvariablen nicht geändert.

Benutzerdefinierte Attribute. Benutzerdefinierte Variablenattribute, die in der Regel mit den Befehl `VARIABLE ATTRIBUTE` der Befehlssyntax erstellt werden.

- Durch Ersetzen werden alle benutzderdefinierten Attribute der Zielvariablen gelöscht und durch die definierten Attribute aus der Quellvariablen ersetzt.
- Durch Zusammenführen werden die definierten Attribute der Quellvariablen mit allen definierten Attributen der Zielvariablen zusammengeführt.

Fehlende Werte. Fehlende Werte sind Werte, die anstelle von fehlenden Werten eingesetzt werden (z. B. 98 für *Weiß nicht* und 99 für *Nicht zutreffend*). In der Regel verfügen auch diese Werte über definierte Wertelabels. Sie beschreiben, welche Bedeutung die Codes für die fehlenden Werte tragen. Alle für die Zielvariable definierten fehlenden Werte werden gelöscht und durch die entsprechenden fehlenden Werte aus der Quellvariablen ersetzt.

Variablenlabel. Aussagekräftige Variablenlabels können Leerzeichen und reservierte Zeichen enthalten, die in Variablennamen nicht zulässig sind. Beim Kopieren von Variableneigenschaften aus einer Quellvariablen in mehrere Zielvariablen sollten Sie sich die Verwendung dieser Option genau überlegen.

Meßniveau. Das Meßniveau kann nominal, ordinal oder metrisch sein. Bei Prozeduren, die zwischen verschiedenen Meßniveaus unterscheiden, werden sowohl nominale als auch ordinale Meßniveaus als **kategorial** betrachtet.

Formate. Bei numerischen Variablen wird über die Formatangabe der numerische Typ (z. B. numerisch, Datum oder Währung), die Breite (Gesamtzahl der angezeigten Zeichen einschließlich der führenden Zeichen, der Abschlußzeichen und des Dezimaltrennzeichens) sowie die Anzahl der angezeigten Dezimalstellen festgelegt. Diese Option gilt nicht für String-Variablen.

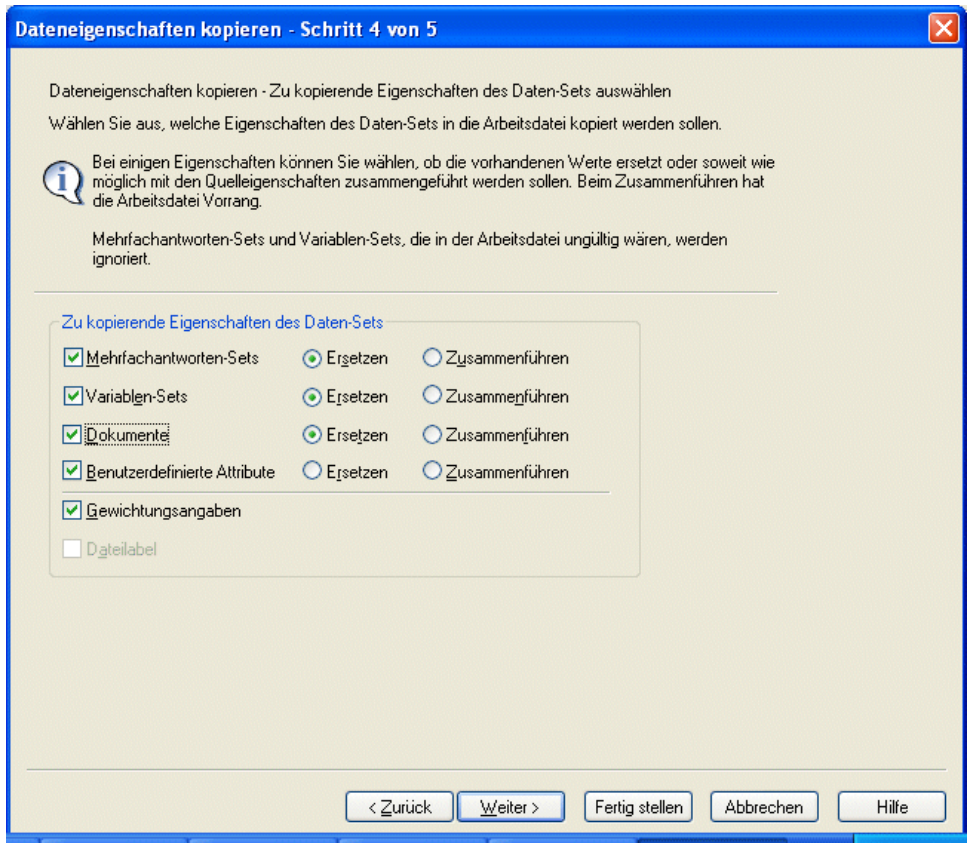
Ausrichtung. Dies betrifft nur die Ausrichtung der Daten in der Datenansicht des Daten-Editors (linksbündig, rechtsbündig, zentriert).

Spaltenbreite im Daten-Editor. Dies betrifft nur die Spaltenbreite in der Datenansicht des Daten-Editors.

Kopieren der (Datei-)Eigenschaften eines Daten-Sets

In einer Quelldatendatei können die globalen Eigenschaften der Daten-Sets ausgewählt und in die Arbeitsdatei übertragen werden. (Diese Option ist nicht verfügbar, wenn die Arbeitsdatei gleichzeitig die Quelldatendatei ist.)

Abbildung 7-10
Assistent zum Kopieren von Dateneigenschaften: Schritt 4


Dateneigenschaften kopieren - Schritt 4 von 5
Dateneigenschaften kopieren - Zu kopierende Eigenschaften des Daten-Sets auswählen
Wählen Sie aus, welche Eigenschaften des Daten-Sets in die Arbeitsdatei kopiert werden sollen.
Bei einigen Eigenschaften können Sie wählen, ob die vorhandenen Werte ersetzt oder soweit wie möglich mit den Quelleigenschaften zusammengeführt werden sollen. Beim Zusammenführen hat die Arbeitsdatei Vorrang.
Mehrfachantworten-Sets und Variablen-Sets, die in der Arbeitsdatei ungültig wären, werden ignoriert.
Zu kopierende Eigenschaften des Daten-Sets
Mehrfachantworten-Sets
Ersetzen
Zusammenführen
Variablen-Sets
Ersetzen
Zusammenführen
Dokumente
Ersetzen
Zusammenführen
Benutzerdefinierte Attribute
Ersetzen
Zusammenführen
Gewichtungsangaben
Dateilabel
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

Mehrfachantworten-Sets. Überträgt Definitionen des Mehrfachantworten-Sets aus der Quelldatendatei in die Arbeitsdatei. (*Anmerkung*: Mehrfachantworten-Sets werden gegenwärtig in der Diagrammerstellung und dem Erweiterungsmodul Tables verwendet.)

- Mehrfachantworten-Sets in der Quelldatendatei, die Variablen ohne Entsprechungen in der Arbeitsdatei enthalten, werden solange ignoriert, bis die entsprechenden Zielvariablen anhand der in Schritt 2 (Auswählen von Quell- und Zielvariablen) angegebenen Einstellungen im Assistenten zum Kopieren von Dateneigenschaften erstellt worden sind.
- Durch Ersetzen werden alle Mehrfachantworten-Sets in der Arbeitsdatei gelöscht und durch die Mehrfachantworten-Sets aus der Quelldatendatei ersetzt.
- Durch Zusammenführen werden den gesamten Mehrfachantworten-Sets in der Arbeitsdatei die Mehrfachantworten-Sets aus der Quelldatendatei hinzugefügt. Wenn beide Dateien ein Set mit demselben Namen enthalten, bleibt das Set in der Arbeitsdatei unverändert.

Variablen-Sets. Mit Variablen-Sets werden die Variablenlisten festgelegt, die in den Dialogfeldern angezeigt werden sollen. Die Definition der Variablen-Sets erfolgt im Menü "Extras" unter "Sets definieren".

- Sets in der Quelldatendatei, die Variablen ohne Entsprechungen in der Arbeitsdatei enthalten, werden solange ignoriert, bis die entsprechenden Zielvariablen anhand der in Schritt 2 (Auswählen von Quell- und Zielvariablen) angegebenen Einstellungen im Assistenten zum Kopieren von Dateneigenschaften erstellt worden sind.
- Durch Ersetzen werden alle Variablen-Sets in der Arbeitsdatei gelöscht und durch Variablen-Sets aus der Quelldatendatei ersetzt.
- Durch Zusammenführen werden den gesamten Variablen-Sets in der Arbeitsdatei die Variablen-Sets aus der Quelldatendatei hinzugefügt. Wenn beide Dateien ein Set mit demselben Namen enthalten, bleibt das Set in der Arbeitsdatei unverändert.

Dokumente. Anmerkungen, die mit dem Befehl `DOCUMENT` an die Datendatei angefügt sind.

- Durch Ersetzen werden alle Dokumente in der Arbeitsdatei gelöscht und durch Dokumente aus der Quelldatendatei ersetzt.
- Durch Zusammenführen werden die Dokumente aus der Quelldatendatei und der Arbeitsdatei kombiniert. Quelldokumente, die in der Arbeitsdatei nicht vorhanden sind, werden der Arbeitsdatei hinzugefügt. Anschließend werden alle Dokumente nach Datum sortiert.

Benutzerdefinierte Attribute. Benutzerdefinierte Datendatei-Attribute, die in der Regel mit dem Befehl `DATAFILE ATTRIBUTE` der Befehlssyntax erstellt werden.

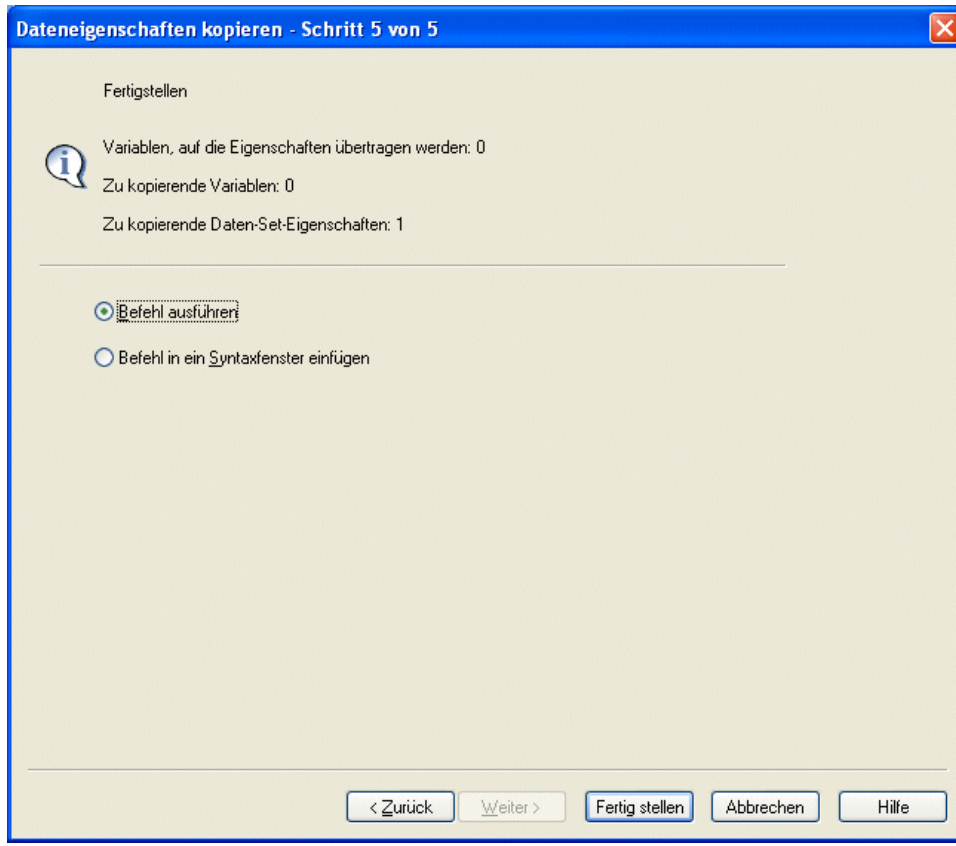
- Durch Ersetzen werden alle vorhandenen benutzerdefinierten Datendatei-Attribute in der Arbeitsdatei gelöscht und durch Datendatei-Attribute aus der Quelldatendatei ersetzt.
- Durch Zusammenführen werden die Datendatei-Attribute aus der Quelldatendatei und der Arbeitsdatei kombiniert. Eindeutige Attributenamen in der Quelldatei, die in der Arbeitsdatei nicht vorhanden sind, werden der Arbeitsdatei hinzugefügt. Wenn derselbe Attributname in beiden Datendateien vorhanden ist, bleibt der Attributname in der Arbeitsdatei unverändert.

Gewichtungsangaben. Mit dieser Option werden Fälle mit der aktuellen Gewichtungsvariablen der Quelldatei gewichtet, sofern in der Arbeitsdatei eine entsprechende Variable vorhanden ist. Dadurch werden alle Gewichtungen überschrieben, die bis dahin in der Arbeitsdatei gültig waren.

Dateilabel. Eine Beschriftung, die einer Datendatei über den Befehl `FILE LABEL` zugeordnet ist.

Ergebnisse

Abbildung 7-11
Assistent zum Kopieren von Dateneigenschaften: Schritt 5



Im letzten Schritt des Assistenten zum Kopieren von Dateieigenschaften erhalten Sie Informationen über die Anzahl der Variablen, für die Variableneigenschaften aus der Quelldatendatei kopiert werden sollen, sowie Informationen über die Anzahl der neu zu erstellenden Variablen und über die Anzahl der Eigenschaften, die aus den Daten-Sets kopiert werden sollen.

Sie können die erstellte Befehlssyntax auch in ein Syntax-Fenster einfügen und sie zur späteren Verwendung speichern.

Ermitteln doppelter Fälle

"Doppelte" Fälle können aus einer Vielzahl von Gründen in Ihren Daten vorkommen. Dazu gehören:

- Dateneingabefehler, bei denen derselbe Fall versehentlich mehrmals eingegeben wurde.
- Mehrere Fälle haben denselben Primär-ID-Wert, aber verschiedene Sekundär-ID-Werte, beispielsweise bei Familienmitgliedern, die alle im selben Haus leben.
- Mehrere Fälle stellen denselben Fall dar, jedoch mit unterschiedlichen Werten für die Variablen, die nicht zur Identifizierung des Falles dienen, beispielsweise mehrere Kaufvorgänge, die von derselben Person oder demselben Unternehmen für verschiedene Produkte oder zu verschiedenen Zeitpunkten durchgeführt wurden.

Mit "Doppelte Fälle ermitteln" haben Sie bei der Definition von *doppelt* sehr große Freiheiten und gewisse Steuerungsmöglichkeiten bei der automatischen Unterscheidung von primären Fällen und doppelten Fällen.

So können Sie doppelte Fälle ermitteln und markieren:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Daten
  Doppelte Fälle ermitteln...

▶ Wählen Sie eine oder mehrere Variablen für die Identifikation übereinstimmender Fälle aus.

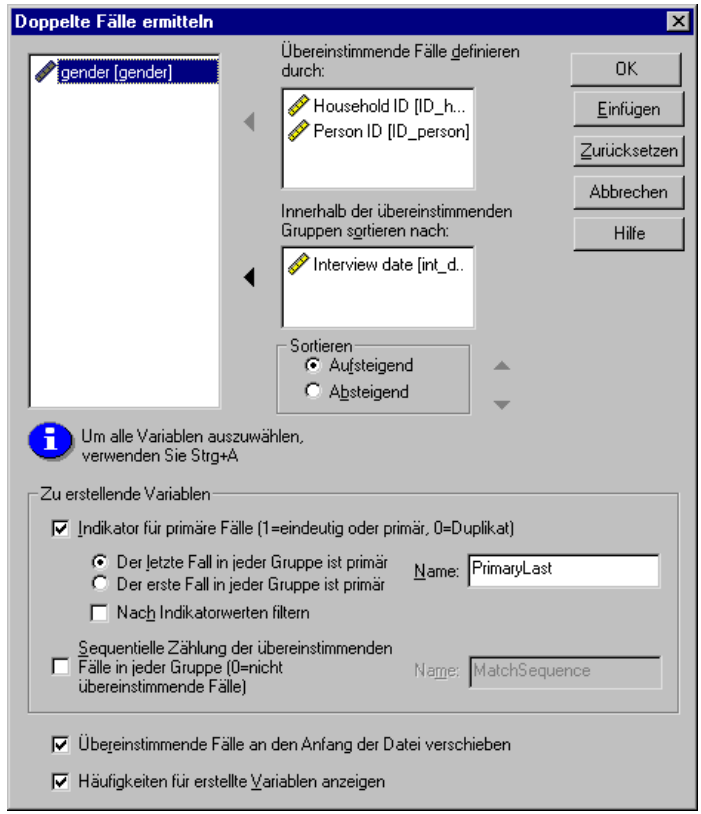
▶ Wählen Sie mindestens eine Option in der Gruppe "Zu erstellende Variablen" aus.

Die folgenden Optionen sind verfügbar:

▶ Auswahl einer oder mehrerer Variablen, um Fälle innerhalb der Gruppen zu sortieren, die durch die ausgewählten Variablen für übereinstimmende Fälle erstellt wurden. Die durch diese Variablen festgelegte Sortierreihenfolge bestimmt den "ersten" und "letzten" Fall in jeder Gruppe. Ansonsten wird die ursprüngliche Dateireihenfolge beibehalten.

▶ Automatisches Filtern doppelter Fälle, so daß sie nicht für Berichte, Diagramme oder statistische Berechnungen verwendet werden.

Abbildung 7-12
Dialogfeld "Doppelte Fälle ermitteln"



Übereinstimmende Fälle definieren durch.Fälle werden als doppelt betrachtet, wenn ihre Werte für *alle* ausgewählten Variablen übereinstimmen. Wenn Sie nur Fälle identifizieren möchten, die in jeder Hinsicht zu 100 % übereinstimmen, müssen Sie alle Variablen auswählen.

Innerhalb der übereinstimmenden Gruppen sortieren nach. Fälle werden automatisch nach den Variablen sortiert, die zur Definition übereinstimmender Fälle dienen. Sie können zusätzliche Sortiervariablen auswählen, die die Reihenfolge der Fälle innerhalb jeder Übereinstimmungsgruppe festlegen.

- Bei jeder Sortiervariablen ist eine Sortierung in aufsteigender und in absteigender Reihenfolge möglich.

- Wenn Sie mehrere Sortiervariablen auswählen, werden die Fälle nach den einzelnen Variablen in den Kategorien der vorhergehenden Variablen in der Liste sortiert. Wenn Sie zum Beispiel *Datum* als erste Sortiervariable und *Menge* als zweite auswählen, wird innerhalb der Datumskategorien nach Menge sortiert.
- Mit den nach oben und nach unten weisenden Pfeil-Schaltflächen rechts neben der Liste können Sie die Sortierreihenfolge der Variablen ändern.
- Die Sortierreihenfolge legt den "ersten" und "letzten" Fall innerhalb jeder Übereinstimmungsgruppe fest, wodurch der Wert der optionalen Indikatorvariablen für primäre Fälle bestimmt wird. Wenn Sie beispielsweise alle außer den aktuellsten Fällen in jeder Übereinstimmungsgruppe herausfiltern möchten, können Sie die Fälle innerhalb der Gruppe in aufsteigender Reihenfolge nach einer Datumsvariablen sortieren, wodurch das aktuellste Datum zum letzten Datum in der Gruppe wird.

Indikator für primäre Fälle. Erstellt eine Variable, die für alle eindeutigen Fälle und den in jeder Gruppe übereinstimmender Fälle als primären Fall identifizierten Fall den Wert 1 und für die nicht-primären doppelten Fälle in jeder Gruppe den Wert 1 annimmt.

- Der primäre Fall kann entweder der letzte oder der erste Fall in jeder Übereinstimmungsgruppe sein. Dies richtet sich nach der Sortierreihenfolge innerhalb der Gruppe. Wenn Sie keine Sortiervariablen angeben, richtet sich die Reihenfolge der Fälle innerhalb der einzelnen Gruppen nach der ursprünglichen Dateireihenfolge.
- Sie können die Indikatorvariable als **Filtervariable** verwenden, um nicht-primäre doppelte Fälle aus Berichten und Analysen auszuschließen, ohne diese Fälle aus der Datendatei zu löschen.

Sequentielle Zählung der übereinstimmenden Fälle in jeder Gruppe. Erstellt eine Variable mit einem Sequenzwert von 1 bis *n* für die Fälle innerhalb der einzelnen Übereinstimmungsgruppen. Die Sequenz beruht auf der aktuellen Reihenfolge der Fälle in jeder Gruppe. Diese ist entweder die ursprüngliche Dateireihenfolge oder die durch angegebene Sortiervariablen festgelegte Reihenfolge.

Übereinstimmende Fälle an den Anfang der Datei verschieben. Sortiert die Datendatei so, daß alle Gruppen übereinstimmender Fälle sich am Anfang der Datendatei befinden. Dadurch wird die visuelle Überprüfung der übereinstimmenden Fälle im Daten-Editor erleichtert.

Häufigkeiten für erstellte Variablen anzeigen. Häufigkeitstabellen mit Zählungen für die einzelnen Werte der erstellten Variablen. Für die Indikatorvariable für primäre Fälle beispielsweise zeigt die Tabelle die Anzahl der Fälle mit dem Wert 0 für diese Variable an, also die Anzahl der doppelten Fälle, und die Anzahl der Fälle mit dem Wert 1 für diese Variable, also die Anzahl der eindeutigen und primären Fälle.

Fehlende Werte. Bei numerischen Variablen wird der systemdefinierte fehlende Wert wie jeder andere Wert behandelt: Fälle mit dem systemdefinierten fehlenden Wert für eine ID-Variable werden so behandelt, als würden sie übereinstimmende Werte für diese Variable aufweisen. Bei String-Variablen werden Fälle ohne Wert für eine ID-Variable so behandelt, als würden sie übereinstimmende Werte für diese Variable aufweisen.

Bereichseinteiler

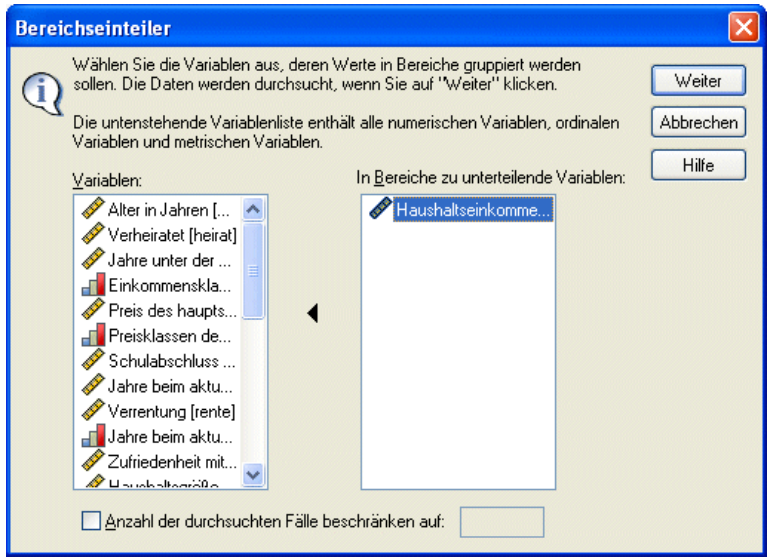
Der Bereichseinteiler soll Sie dabei unterstützen, neue Variablen zu erstellen. Als Grundlage dafür werden nebeneinanderliegende Werte bestehender Variablen in eine begrenzte Anzahl unterschiedlicher Kategorien gruppiert. Sie können den Bereichseinteiler für folgende Vorgänge verwenden:

- Erstellen kategorialer Variablen aus nebeneinanderliegenden metrischen Variablen. Sie können beispielsweise eine metrische Variable für das Einkommen verwenden, um eine neue kategoriale Variable zu erstellen, die Einkommensbereiche enthält.
- Reduzieren einer großen Zahl ordinaler Kategorien zu einer kleineren Menge von Kategorien. Sie können beispielsweise eine neunstufige Bewertungsskala auf drei Kategorien, niedrig, mittel und hoch, reduzieren.

Zu Beginn der Verwendung des Bereichseinteilers führen Sie folgenden Schritt durch:

▶ Wählen Sie die numerischen metrischen und/oder ordinalen Variablen aus, für die neue kategoriale (in Bereiche eingeteilte) Variablen erstellt werden sollen.

Abbildung 7-13
Erstes Dialogfeld für die Auswahl der in Bereiche einzuteilenden Variablen



Wahlweise können Sie die Anzahl der zu durchsuchenden Fälle begrenzen. Bei Datendateien mit einer großen Anzahl an Fällen kann die Begrenzung der zu durchsuchenden Fälle Zeit sparen, doch Sie sollten dies nach Möglichkeit vermeiden, da es die Verteilung der Werte beeinflußt, die in den darauffolgenden Berechnungen im Bereichseinteiler verwendet werden.

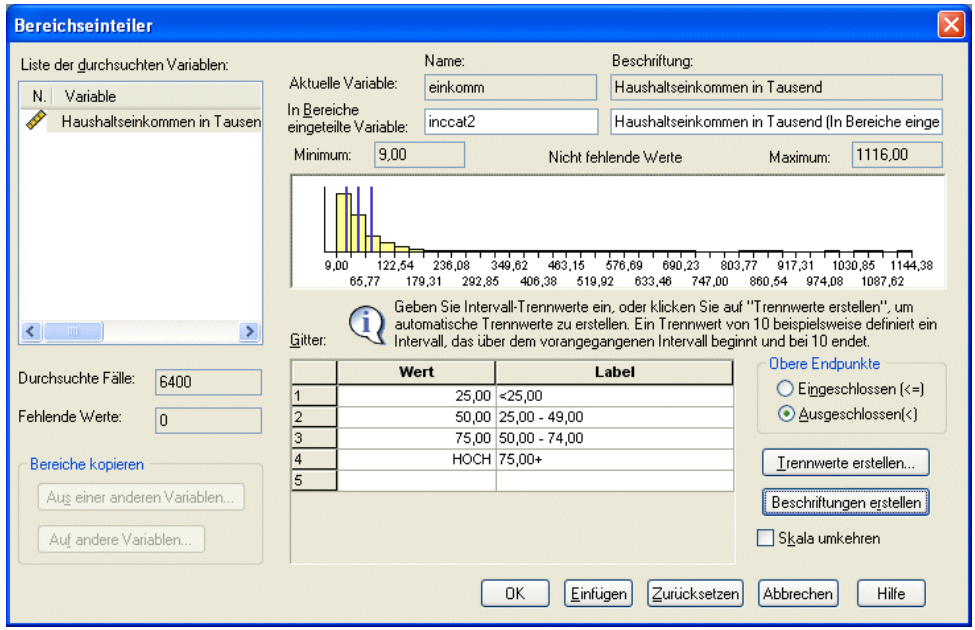
Anmerkung: String-Variablen und nominale numerische Variablen werden nicht in der Liste der Quellvariablen angezeigt. Für den Bereichseinteiler sind numerische Variablen erforderlich, die entweder auf einem metrischen oder einem ordinalen Meßniveau gemessen wurden, da er davon ausgeht, daß die Datenwerte eine logische Reichenfolge aufweisen, die für eine sinnvolle Gruppierung der Werte verwendet werden kann. In der Variablenansicht des Daten-Editors kann das definierte Messniveau einer Variablen geändert werden. Für weitere Informationen siehe “Meßniveau einer Variablen” in Kapitel 5 auf S. 99.

So teilen Sie Variablen in Bereiche ein:

- Wählen Sie die folgenden Befehle aus den Menüs im Fenster "Daten-Editor" aus:
  Transformieren
  Bereichseinteiler...
- Wählen Sie die numerischen metrischen und/oder ordinalen Variablen aus, für die neue kategoriale (in Bereiche eingeteilte) Variablen erstellt werden sollen.
- Wählen Sie eine Variable in der Liste der durchsuchten Variablen aus.
- Geben Sie einen Namen für die neue in Bereiche eingeteilte Variable ein. Variablennamen müssen eindeutig sein und den Regeln für Variablennamen von SPSS entsprechen. Für weitere Informationen siehe "Variablennamen" in Kapitel 5 auf S. 99.
- Definieren Sie die Bereichseinteilungskriterien für die neue Variable. Für weitere Informationen siehe "Einteilen von Variablen in Bereiche" auf S. 155.
- Klicken Sie auf OK.

Einteilen von Variablen in Bereiche

Abbildung 7-14
Bereichseinteiler, Hauptdialogfeld



Das Hauptdialogfeld des Bereichseinteilers enthält folgende Informationen über die durchsuchten Variablen:

Liste der durchsuchten Variablen. Zeigt die Variablen an, die Sie im ersten Dialogfeld ausgewählt haben. Sie können die Liste anhand des Meßniveaus (metrisch oder ordinal) oder anhand des Variablenlabels oder -namens sortieren, indem Sie auf die Spaltenüberschriften klicken.

Durchsuchte Fälle. Gibt die Zahl der durchsuchten Fälle an. Alle durchsuchten Fälle ohne systemdefinierte oder benutzerdefinierte fehlende Werte für die ausgewählte Variable werden verwendet, um die in den Berechnungen des Bereichseinteilers verwendete Werteverteilung zu erstellen. Dazu gehören auch das im Hauptdialogfeld angezeigte Histogramm und Trennwerte auf der Grundlage von Perzentilen oder Einheiten der Standardabweichung.

Fehlende Werte. Gibt die Anzahl der durchsuchten Fälle mit systemdefinierten oder benutzerdefinierten fehlenden Werten an. Fehlende Werte werden in keiner der in Bereiche eingeteilten Kategorien verwendet. Für weitere Informationen siehe "Benutzerdefinierte fehlende Werte im Bereichseinteiler" auf S. 163.

Aktuelle Variable. Der Name und das Variablenlabel (sofern vorhanden) für die derzeit ausgewählte Variable, die als Grundlage für die neue, in Bereiche eingeteilte Variable dient.

In Bereiche eingeteilte Variable. Der Name und gegebenenfalls das Variablenlabel für die neue, in Bereiche eingeteilte Variable.

- **Name.** Sie müssen einen Namen für die neue Variable eingeben. Variablennamen müssen eindeutig sein und den Regeln für Variablennamen von SPSS entsprechen. Für weitere Informationen siehe "Variablennamen" in Kapitel 5 auf S. 99.
- **Label (Beschriftung).** Sie können ein aussagekräftiges Variablenlabel mit bis zu 255 Zeichen eingeben. Das Standard-Variablenlabel ist das Variablenlabel (sofern vorhanden) oder der Variablenname der Quellvariable, wobei am Ende des Labels *(In Bereiche eingeteilt)* angehängt ist.

Minimum und Maximum. Der Mindest- und Höchstwert für die derzeit ausgewählte Variable, auf der Grundlage der durchsuchten Fälle ohne die Werte, die als benutzerdefiniert fehlend definiert wurden.

Nichtfehlende Werte. Das Histogramm zeigt die Verteilung der nichtfehlenden Werte für die derzeit ausgewählte Variable (auf der Grundlage der durchsuchten Fälle) an.

- Nach der Definition von Bereichen für die neue Variable, werden im Histogramm vertikale Linien angezeigt, um die Trennwerte für die Bereichsdefinition anzuzeigen.
- Sie können auf die Trennwertlinien klicken und sie an andere Stellen im Histogramm ziehen, um so die Größe der Bereiche zu verändern.
- Sie können Bereiche entfernen, indem Sie die Trennwertlinien vom Histogramm weg ziehen.

Anmerkung: Das Histogramm (mit den nichtfehlenden Werten), das Minimum und das Maximum beruhen auf den durchsuchten Werten. Wenn Sie nicht alle Fälle durchsuchen lassen, wird die tatsächliche Verteilung möglicherweise nicht richtig wiedergegeben, insbesondere, wenn die Datendatei anhand der ausgewählten

Variablen sortiert wurde. Wenn Sie 0 Fälle durchsuchen, stehen keine Informationen über die Werteverteilung zur Verfügung.

Gitter. Zeigt die Werte an, die die oberen Endpunkte der einzelnen Bereiche darstellen, sowie gegebenenfalls die Wertelabels für die einzelnen Bereiche.

- **Wert.** Die Werte, die die oberen Endpunkte der einzelnen Bereiche darstellen. Sie können Werte eingeben oder mit Hilfe von Trennwerte erstellen Bereiche automatisch anhand ausgewählter Kriterien erstellen. Standardmäßig werden Trennwerte mit dem Wert *HOCH* automatisch aufgenommen. Dieser Bereich enthält alle nichtfehlenden Werte, die über den anderen Trennwerten liegen. Der durch den untersten Trennwert definierte Bereich enthält alle nichtfehlenden Werte kleiner oder gleich diesem Wert (oder nur kleiner als dieser Wert, je nachdem, wie Sie die oberen Endpunkte definieren).
- **Label (Beschriftung).** Optionale, aussagekräftige Labels für die Werte der neuen, in Bereiche eingeteilten Variablen. Da die Werte der neuen Variablen einfach aufeinanderfolgende Ganzzahlen von 1 bis *n* sind, können Labels, die angeben, wofür die Werte stehen, sehr hilfreich sein. Sie können Labels eingeben oder mit Hilfe von Beschriftungen erstellen automatisch Wertelabels erstellen.

So löschen Sie einen Bereich aus dem Gitter

▶ Klicken Sie mit der rechten Maustaste entweder auf die Zelle *Wert* oder auf die Zelle *Label (Beschriftung)* für den Bereich.

▶ Wählen Sie im Kontextmenü die Option Zeile löschen.

Anmerkung: Wenn Sie den Bereich *HOCH* wählen, wird allen Fällen mit Werten, die höher sind als der letzte angegebene Trennwert, bei der neuen Variablen der Wert "Systemdefiniert fehlend" zugewiesen.

So löschen Sie alle Labels bzw. alle definierten Bereiche

▶ Klicken Sie mit der rechten Maustaste auf eine beliebige Stelle im Gitter.

▶ Wählen Sie im Kontextmenü entweder die Option Alle Beschriftungen löschen oder die Option Alle Trennwerte löschen.

Obere Endpunkte. Hiermit wird die Behandlung der Werte für die oberen Endpunkte in der Spalte *Wert* des Gitters festgelegt.

- **Eingeschlossen (<=).** Fälle mit dem in der Zelle *Wert* angegebenen Wert werden in die in Bereiche eingeteilte Kategorie aufgenommen. Wenn Sie beispielsweise die Werte 25, 50 und 75 angeben, werden Fälle mit einem Wert von exakt 25 in den ersten Bereich eingeordnet, da dieser alle Fälle mit Werten kleiner oder gleich 25 enthält.
- **Ausgeschlossen (<).** Fälle mit dem in der Zelle *Wert* angegebenen Wert werden nicht in die in Bereiche eingeteilte Kategorie aufgenommen. Statt dessen werden sie in den nächsten Bereich aufgenommen. Wenn Sie beispielsweise die Werte 25, 50 und 75 angeben, werden Fälle mit einem Wert von exakt 25 in den zweiten und nicht in den ersten Bereich eingeordnet, da der erste Bereich nur Fälle mit Werten kleiner als 25 enthält.

Trennwerte erstellen. Erstellt automatisch in Bereiche eingeteilte Kategorien für Intervalle mit gleicher Breite, Intervalle mit derselben Anzahl von Fällen oder auf Standardabweichungen beruhende Intervalle. Diese Option ist nicht verfügbar, wenn 0 Fälle durchsucht wurden. Für weitere Informationen siehe “Automatisches Erstellen von in Bereiche eingeteilten Kategorien” auf S. 158.

Beschriftungen erstellen. Erstellt aussagekräftige Beschreibungen für die sequentiellen ganzzahligen Werte der neuen in Bereiche eingeteilten Variablen, und zwar auf der Grundlage der Werte im Gitter und der angegebenen Behandlung der oberen Endpunkte (eingeschlossen oder ausgeschlossen).

Skala umkehren. Standardmäßig sind die Werte der neuen, in Bereiche eingeteilten Variablen aufsteigende sequentielle Ganzzahlen von 1 bis *n*. Durch Umkehr der Skala werden die Werte zu absteigenden sequentiellen Ganzzahlen von *n* bis 1.

Bereiche kopieren. Sie können die Spezifikationen für die Bereichseinteilung von einer anderen Variablen auf die derzeit ausgewählte Variable oder von der ausgewählten Variable auf mehrere andere Variablen kopieren. Für weitere Informationen siehe “Kopieren von in Bereiche eingeteilten Kategorien” auf S. 161.

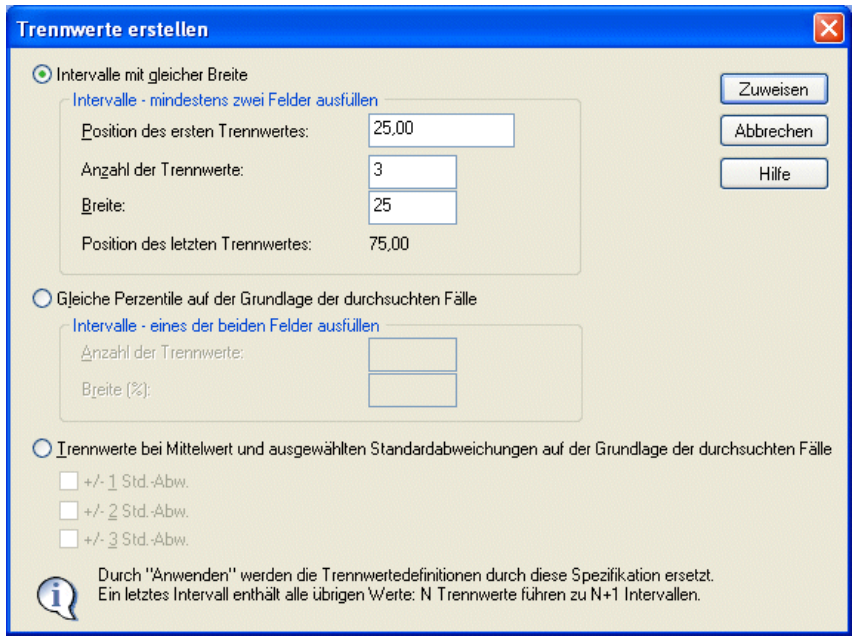
Automatisches Erstellen von in Bereiche eingeteilten Kategorien

Im Dialogfeld “Trennwerte erstellen” können Sie automatisch in Bereiche eingeteilte Kategorien erstellen, die auf den ausgewählten Kriterien beruhen.

So verwenden Sie das Dialogfeld "Trennwerte erstellen"

- Wählen Sie (durch Klicken) eine Variable in der Liste der durchsuchten Variablen aus.
- Klicken Sie auf Trennwerte erstellen.
- Wählen Sie die Kriterien für das Erstellen der Trennwerte aus, welche die in Bereiche eingeteilten Kategorien definieren.
- Klicken Sie auf Zuweisen.

Abbildung 7-15
Dialogfeld "Trennwerte erstellen"



Anmerkung: Das Dialogfeld "Trennwerte erstellen" ist nicht verfügbar, wenn 0 Fälle durchsucht wurden.

Intervalle mit gleicher Breite. Erstellt in Bereiche eingeteilte Kategorien mit gleicher Breite (z. B. 1-10, 11-20, 21-30 usw.), die auf zwei der folgenden drei Kriterien beruhen:

- **Position des ersten Trennwertes.** Der Wert, der das obere Ende der untersten Kategorie für die Bereichseinteilung kennzeichnet (Beispiel: Der Wert 10 gibt einen Bereich an, der alle Werte bis 10 einschließt).
- **Anzahl der Trennwerte.** Die Anzahl der in Bereiche eingeteilten Kategorien ist die Anzahl der Trennwerte *plus 1*. So führen 9 Trennwerte zu 10 in Bereiche eingeteilte Kategorien.
- **Stärke.** Die Breite der einzelnen Intervalle. Der Wert 10 beispielsweise teilt *Alter in Jahren* in jeweils 10 Jahre umfassende Intervalle ein.

Gleiche Perzentile auf der Grundlage der durchsuchten Fälle. Erstellt in Bereiche eingeteilte Kategorien mit der gleichen Anzahl von Fällen in jedem Bereich (unter Verwendung des empirischen Algorithmus für Perzentile). Als Grundlage dient eines der folgenden Kriterien:

- **Anzahl der Trennwerte.** Die Anzahl der in Bereiche eingeteilten Kategorien ist die Anzahl der Trennwerte *plus 1*. So führen drei Trennwerte zu vier Perzentilbereichen (Quartilen) mit jeweils 25 % der Fälle.
- **Breite (%).** Die Breite der einzelnen Intervalle als Prozentsatz der Gesamtanzahl der Fälle. Der Wert 33,3 beispielsweise führt zu drei in Bereiche eingeteilten Kategorien (zwei Trennwerte) mit jeweils 33,3 % der Fälle.

Wenn die Quellvariable eine relativ geringe Anzahl eindeutiger Werte oder eine große Anzahl von Fällen mit demselben Wert enthält, erhalten Sie möglicherweise weniger Bereiche als angefordert. Liegen mehrere identische Werte an einem Trennwert vor, werden alle Werte in dasselbe Intervall aufgenommen. Die tatsächlichen Prozentsätze sind daher nicht in jedem Fall genau gleich.

Trennwerte bei Mittelwert und ausgewählten Standardabweichungen auf der Grundlage der durchsuchten Fälle Erstellt in Bereiche eingeteilte Kategorien auf der Grundlage der Werte für Mittelwert und Standardabweichung für die Verteilung der Variablen.

- Wenn Sie keines der Standardabweichungs-Intervalle auswählen, werden zwei in Bereiche eingeteilte Kategorien erstellt, mit dem Mittelwert als Trennwert zwischen den Bereichen.
- Sie können eine beliebige Kombination von Standardabweichungs-Intervallen auf der Grundlage von einer, zwei und/oder drei Standardabweichungen auswählen. Beispiel: Wenn Sie alle drei Möglichkeiten auswählen, würde das zu 8 in Bereiche eingeteilten Kategorien führen – 6 Bereiche in einem Standardabweichungs-Intervall und zwei Bereiche für Fälle, die mehr als drei Standardabweichungen über bzw. unter dem Mittelwert liegen.

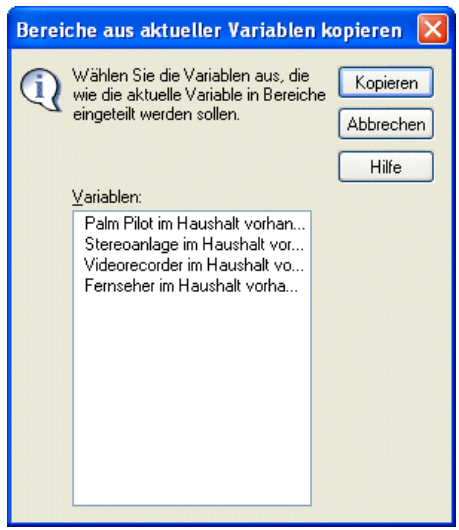
Bei einer Normalverteilung liegen 68 % der Fälle innerhalb einer Standardabweichung vom Mittelwert, 95 % innerhalb von zwei Standardabweichungen und 99 % innerhalb von drei Standardabweichungen. Das Erstellen von in Bereichen eingeteilten Kategorien auf der Grundlage von Standardabweichungen kann zu definierten Bereichen außerhalb des tatsächlichen Datenbereichs und sogar außerhalb des Bereichs der möglichen Datenwerte (z. B. ein negativer Gehaltsbereich) führen.

Anmerkung: Die Berechnung von Perzentilen und Standardabweichungen beruht auf den durchsuchten Fällen. Wenn Sie die Anzahl der durchsuchten Fälle beschränken, enthalten die resultierenden Bereiche möglicherweise nicht den Anteil an Fällen, den Sie in diesen Bereichen haben wollten, insbesondere dann, wenn die Datendatei nach der Quellvariable sortiert wird. Beispiel: Wenn Sie nur die ersten 100 Fälle einer Datendatei mit 1000 Fällen durchsuchen lassen und die Datendatei in aufsteigender Reihenfolge nach dem Alter des Befragten sortiert ist, erhalten Sie möglicherweise nicht vier Perzentil-Altersbereichen mit jeweils 25 % der Fälle, sondern in den ersten drei Fällen befinden sich vielleicht nur jeweils 3,3 % der Fälle und im letzten Bereich 90 %.

Kopieren von in Bereiche eingeteilten Kategorien

Beim Erstellen von in Bereiche eingeteilten Kategorien für mehrere Variablen, können Sie die Spezifikationen für die Bereichseinteilung von einer anderen Variablen auf die derzeit ausgewählte Variable oder von der ausgewählten Variablen auf mehrere andere Variablen kopieren.

Abbildung 7-16
Kopieren von Bereichen von der aktuellen Variablen bzw. auf die aktuelle Variable


Bereiche aus aktueller Variablen kopieren
Wählen Sie die Variablen aus, die wie die aktuelle Variable in Bereiche eingeteilt werden sollen.
Kopieren
Abbrechen
Hilfe
Variablen:
Palm Pilot im Haushalt vorhan...
Stereoanlage im Haushalt vor...
Videorecorder im Haushalt vo...
Fernseher im Haushalt vorha...

So kopieren Sie Bereichseinteilungsspezifikationen

- Definieren Sie in Bereiche eingeteilte Kategorien für mindestens eine Variable – klicken Sie jedoch *nicht* auf OK oder Einfügen.
- Wählen Sie (durch Klicken) eine Variable in der Liste der durchsuchten Variablen aus, für die Sie in Bereiche eingeteilte Kategorien definiert haben.
- Klicken Sie auf Auf andere Variablen.
- Wählen Sie die Variablen aus, für die neue Variablen mit denselben in Bereiche eingeteilten Kategorien erstellt werden sollen.
- Klicken Sie auf Kopieren.

  oder

- Wählen Sie (durch Klicken) eine Variable in der Liste der durchsuchten Variablen aus, auf die Sie in Bereiche eingeteilte Kategorien kopieren möchten.
- Klicken Sie auf Aus einer anderen Variablen.

- Wählen Sie die Variable mit den in Bereiche eingeteilten Kategorien aus, die Sie kopieren möchten.
- Klicken Sie auf Kopieren.

Wenn Sie Wertelabels für die Variable angegeben haben, aus der Sie die Bereichseinteilungsspezifikationen kopieren, werden diese ebenfalls kopiert.

Anmerkung: Wenn Sie im Hauptdialogfeld des Bereichseinteilers auf OK klicken, um die neuen in Bereiche eingeteilten Variablen zu erstellen (oder das Dialogfeld auf andere Weise schließen), können Sie den Bereichseinteiler nicht dazu verwenden, diese in Bereiche eingeteilten Kategorien auf andere Variablen zu verschieben.

Benutzerdefinierte fehlende Werte im Bereichseinteiler

Als benutzerdefiniert fehlend definierte Werte (Werte, die als Codes für fehlende Daten gekennzeichnet wurden) für die Quellvariable werden nicht in die in Bereiche eingeteilten Kategorien für die neue Variable aufgenommen. Benutzerdefinierte fehlende Werte für die Quellvariable werden als benutzerdefinierte fehlende Werte für die neue Variable kopiert, und alle definierten Wertelabels für die Codes für fehlende Werte werden ebenfalls kopiert.

Wenn ein Code für einen fehlenden Wert mit einem der Werte für in Bereiche eingeteilte Kategorien für die neue Variable in Konflikt steht, wird der Code für den fehlenden Wert für die neue Variable als nicht in Konflikt stehender Wert aufgezeichnet, indem zum höchsten Wert für eine in Bereiche eingeteilte Kategorie der Wert 100 addiert wird. Beispiel: Wenn der Wert 1 für die Quellvariable als benutzerdefiniert fehlend definiert ist und die neue Variable sechs in Bereiche eingeteilte Kategorien umfaßt, haben alle Fälle mit dem Wert 1 für die neue Quellvariable den Wert 106, und 106 wird als benutzerdefiniert fehlend definiert. Wenn für den benutzerdefinierten fehlenden Wert für die Quellvariable ein Wertelabel definiert war, wird dieses Label als Wertelabel für den aufgezeichneten Wert der neuen Variablen beibehalten.

Anmerkung: Wenn die Quellvariable einen definierten Bereich benutzerdefinierter fehlender Werte der Form *LO-n* enthält (wobei *n* eine positive Zahl ist), sind die entsprechenden benutzerdefinierten fehlenden Werte für die neue Variable negative Zahlen.

Kapitel
8

Transformieren von Daten

Im Idealfall sind Ihre Rohdaten genau für die Analyse geeignet, die Sie ausführen möchten, und die Beziehungen zwischen den Variablen sind entweder linear oder rein orthogonal. Dies ist leider selten der Fall. Mit einer Vorabanalyse können problematische Kodierschemata oder Kodierfehler erkannt werden. Transformationen von Daten können auch erforderlich sein, um die tatsächliche Beziehung zwischen den Variablen herauszuarbeiten.

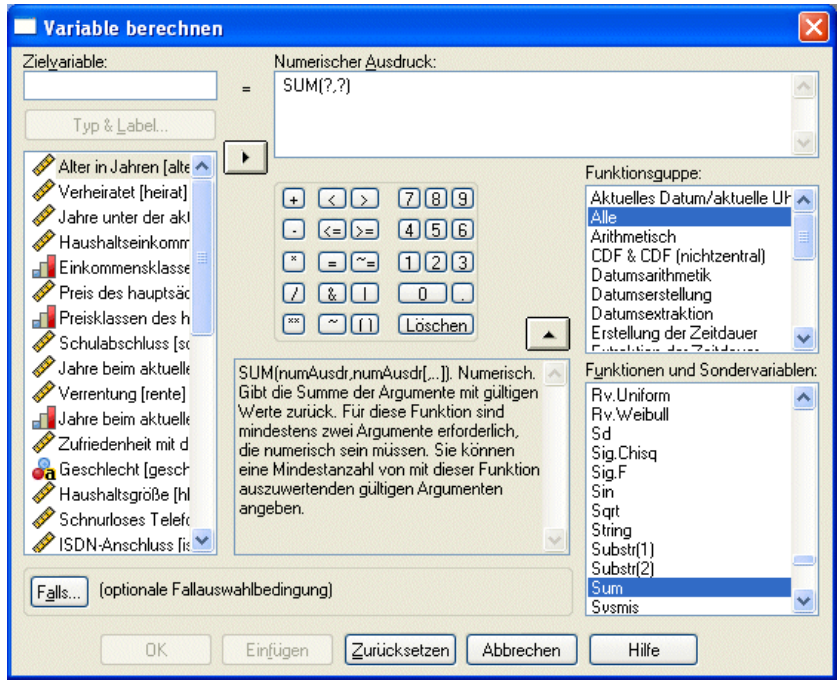
Mit SPSS können Sie verschiedene Transformationen von Daten ausführen, von einfachen Aufgaben wie dem Zusammenfassen von Kategorien zur Analyse, bis zu fortgeschritteneren Aufgaben wie dem Erstellen neuer Variablen auf der Grundlage von Bedingungen und komplizierten Gleichungen.

Berechnen von Variablen

Im Dialogfeld “Berechnen” werden Werte für Variablen auf der Grundlage von numerischen Transformationen anderer Variablen berechnet.

- Sie können Werte für numerische oder String-Variablen berechnen.
- Sie können neue Variablen erstellen oder die Werte vorhandener Variablen ersetzen. Bei neuen Variablen können Sie außerdem Variablentyp und -label angeben.
- Auf der Grundlage von logischen Bedingungen können Sie Werte für ausgewählte Teilmengen von Daten berechnen lassen.
- Sie können über 70 systemeigene Funktionen verwenden, darunter arithmetische Funktionen, Statistikfunktionen, Verteilungsfunktionen und String-Funktionen.

Abbildung 8-1
Dialogfeld "Variable berechnen"



So berechnen Sie Variablen:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Transformieren
  Berechnen...

- Geben Sie den Namen einer einzelnen Zielvariablen ein. Dies kann eine vorhandene Variable sein oder eine neue Variable, die in die Arbeitsdatei aufgenommen werden soll.

- Um einen Ausdruck zu bilden, fügen Sie die Elemente in das Feld "Ausdruck" ein, oder geben Sie den Ausdruck direkt in dieses Feld ein.

  - Sie können Funktionen oder häufig verwendete Systemvariablen einfügen, indem Sie eine Gruppe aus der Liste "Funktionsgruppe" auswählen und in der Liste "Funktionen und Sodervariablen" auf die Funktion bzw. Variable doppelklicken (oder die Funktion bzw. Variable auswählen und auf den Pfeil

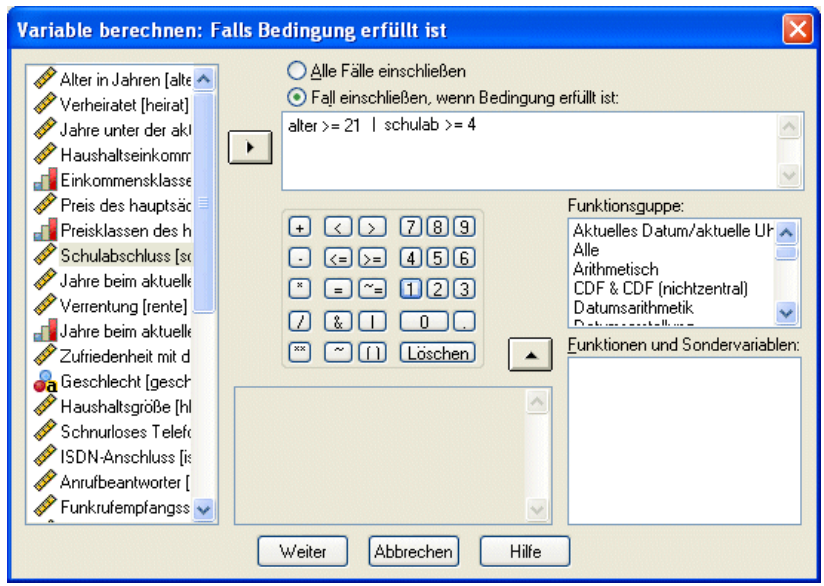
neben der Liste “Funktionsgruppe” klicken). Geben Sie alle durch Fragezeichen gekennzeichneten Parameter an (gilt nur für Funktionen). Die Funktionsgruppe mit der Beschriftung Alle bietet eine Auflistung aller verfügbaren Funktionen und Systemvariablen. Eine kurze Beschreibung der aktuell ausgewählten Funktion oder Variablen wird in einem speziellen Bereich des Dialogfelds angezeigt.

- String-Konstanten müssen in Anführungszeichen oder Apostrophe eingeschlossen werden.
- Wenn die Werte Dezimalstellen enthalten, muss ein Punkt (.) als Dezimaltrennzeichen verwendet werden.
- Bei neuen String-Variablen müssen Sie außerdem Typ & Label auswählen, um den Datentyp anzugeben.

Variable berechnen: Falls Bedingung erfüllt ist

Im Dialogfeld “Variable berechnen: Falls Bedingung erfüllt ist” können Sie Transformationen auf Teilmengen von Fällen anwenden, die anhand bedingter Ausdrücke ausgewählt werden. Ein bedingter Ausdruck gibt für jeden Fall den Wert *Wahr*, *Falsch* oder *Fehlend* zurück.

Abbildung 8-2
Dialogfeld "Variable berechnen: Falls Bedingung erfüllt ist"



- Wenn das Ergebnis eines bedingten Ausdrucks *Wahr* ist, wird der Fall in die ausgewählte Untergruppe aufgenommen.
- Wenn das Ergebnis eines bedingten Ausdrucks *Falsch* oder *Fehlend* ist, wird der Fall nicht in die ausgewählte Untergruppe aufgenommen.
- In den meisten bedingten Ausdrücke wird mindestens einer der sechs Vergleichsoperatoren (<, >, <=, >=, = und ~=) verwendet. Diese sind auf der Rechentastatur verfügbar.
- Bedingte Ausdrücke können Variablennamen, Konstanten, arithmetische Operatoren, numerische (und andere) Funktionen, logische Variablen und Vergleichsoperatoren enthalten.

Variable berechnen: Typ und Label

In der Standardeinstellung sind die neu berechneten Variablen numerisch. Zum Berechnen einer neuen String-Variablen müssen Sie Datentyp und Länge angeben.

Label (Beschriftung). Optionale, aussagekräftige Variablenlabels können bis zu 120 Zeichen umfassen. Sie können ein Label eingeben oder die ersten 110 Zeichen des berechneten Ausdrucks als Label verwenden.

Typ. Es können numerische oder String-Variablen (alphanumerische Variablen) berechnet werden. In Berechnungen können keine String-Variablen verwendet werden.

Abbildung 8-3
Dialogfeld "Variable berechnen: Typ und Label"

Funktionen

Es werden verschiedene Typen von Funktionen unterstützt. Dazu gehören:

- Arithmetische Funktionen
- Statistische Funktionen
- String-Funktionen
- Datums- und Uhrzeitfunktionen,
- Verteilungsfunktionen
- Funktionen mit Zufallsvariablen,
- Funktionen mit fehlenden Werten.
- Bewertungsfunktionen (nur SPSS-Server)

Eine vollständige Liste der Funktionen von SPSS erhalten Sie, indem Sie im Index der Online-Hilfe nach Funktionen suchen.

Fehlende Werte in Funktionen

Fehlende Werte werden von Funktionen und einfachen arithmetischen Ausdrücken unterschiedlich behandelt. In dem Ausdruck:

```
(var1+var2+var3)/3
```

fehlt das Ergebnis, wenn ein Fall einen fehlenden Wert für eine der drei Variablen enthält.

In dem Ausdruck:

```
MEAN(var1,var2,var3)
```

fehlt das Ergebnis nur, wenn der Fall fehlende Werte für alle drei Variablen enthält.

Bei statistischen Funktionen können Sie die Mindestanzahl von Argumenten angeben, die nichtfehlende Werte enthalten müssen. Geben Sie dazu nach dem Namen der Funktion einen Punkt und die Mindestanzahl ein, wie zum Beispiel in

```
MEAN.2(var1,var2,var3)
```

Zufallszahlengeneratoren

Im Dialogfeld "Zufallszahlengeneratoren" können Sie den Zufallszahlengenerator auswählen und den Startwert für eine Sequenz von Zufallszahlen festlegen.

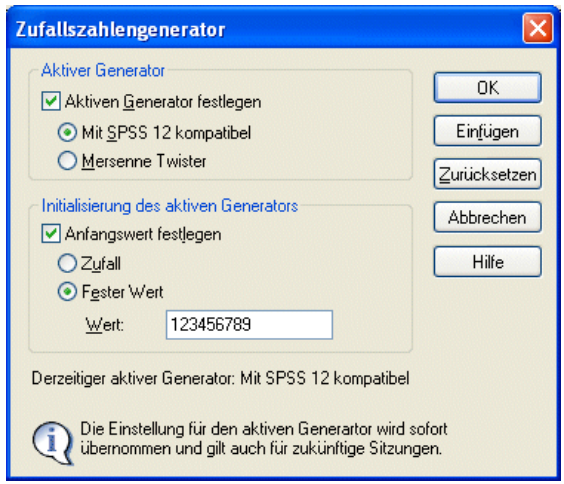
Aktiver Generator. Zwei verschiedene Zufallszahlengeneratoren stehen zur Verfügung:

- **SPSS12-kompatibel.** Der Zufallszahlengenerator, der in SPSS 12 und älteren Versionen verwendet wird. Wenn zufallsbestimmte Ergebnisse reproduziert werden sollen, die in einer älteren Version mit einem bestimmten Startwert erzeugt wurden, ist dieser Zufallszahlengenerator zu verwenden.
- **Mersenne-Twister.** Ein neuerer Zufallszahlengenerator, der für Simulationszwecke eine höhere Zuverlässigkeit bietet. Sofern es nicht darum geht, zufallsbestimmte Ergebnisse aus SPSS 12 oder älteren Versionen zu reproduzieren, sollte man diesen Zufallszahlengenerator verwenden.

Initialisierung des aktiven Generators. Der Startwert für Zufallszahlen ändert sich jedesmal, wenn SPSS eine Zufallszahl zur Verwendung bei Transformationen (wie bei den Zufallsdistributionsfunktionen), der Entnahme von Zufallsstichproben oder der Gewichtung von Fällen erzeugt. Wenn Sie eine Reihe von Zufallszahlen

reproduzieren möchten, setzen Sie den Startwert auf einen bestimmten Wert zurück, bevor eine Analyse mit diesen Zufallszahlen ausgeführt wird. Dieser Wert muß eine positive Ganzzahl sein.

Abbildung 8-4
Dialogfeld "Zufallszahlengenerator"



So können Sie den Zufallszahlengenerator auswählen und/oder den Startwert festlegen:

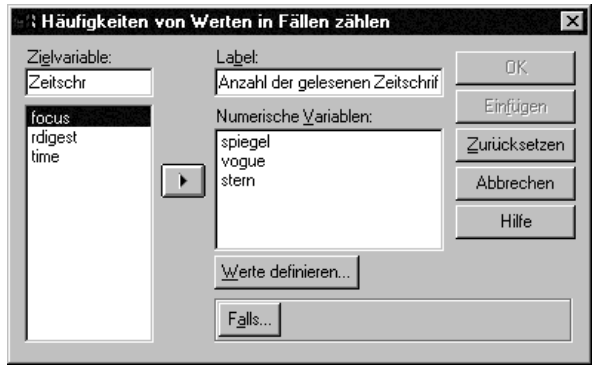
▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Transformieren
  Zufallszahlengeneratoren

Häufigkeiten von Werten in Fällen zählen

In diesem Dialogfeld wird eine Variable erstellt, mit welcher das Auftreten derselben Werte in einer Variablenliste pro Fall gezählt wird. Zum Beispiel könnte eine Befragung eine Liste von Zeitschriften mit Feldern zum Ankreuzen für *Ja* und *Nein* enthalten, mit denen die Befragten angeben, welche Zeitschriften sie lesen. Sie könnten dann die Anzahl aller Antworten mit *Ja* für jeden Befragten zählen und eine neue Variable erstellen, welche die Anzahl der gelesenen Zeitschriften enthält.

Abbildung 8-5
Dialogfeld "Häufigkeiten von Werten in Fällen zählen"



So zählen Sie die Werte in Fällen

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Transformieren
  Zählen…

- Geben Sie einen Namen für die Zielvariable ein.
- Wählen Sie mindesten zwei Variablen desselben Typs aus (numerische oder String-Variablen).
- Klicken Sie auf Werte definieren, und geben Sie an, welcher Wert oder welche Werte gezählt werden sollen.

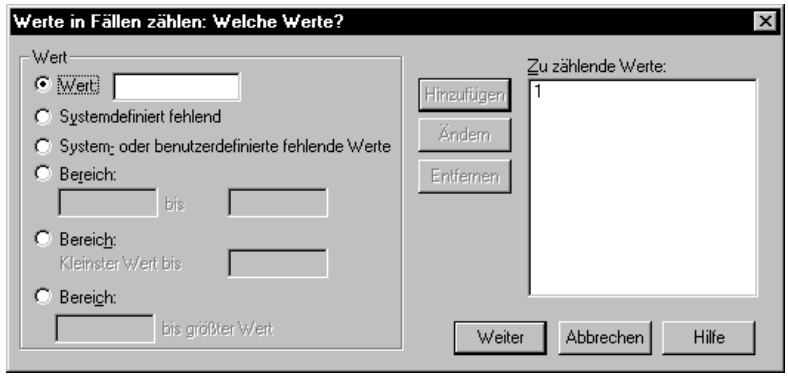
  Wahlweise können Sie eine Teilmenge von Fällen definieren, für die Werte gezählt werden sollen.

Werte in Fällen zählen: Welche Werte?

Der Wert der Zielvariablen (im Hauptdialogfeld) wird jedesmal um 1 erhöht, wenn eine der ausgewählten Variablen einer Angabe in der Liste "Zu zählende Werte" entspricht. Wenn ein Fall auf mehrere Angaben für eine Variable zutrifft, wird die Zielvariable für diese Variablen mehrmals erhöht.

Angaben von Werten können einzelne Werte, fehlende oder systemdefinierte fehlende Werte und Bereiche enthalten. Bei Bereichen sind die Endwerte und alle benutzerdefinierten fehlenden Werte eingeschlossen, die in den Bereich fallen.

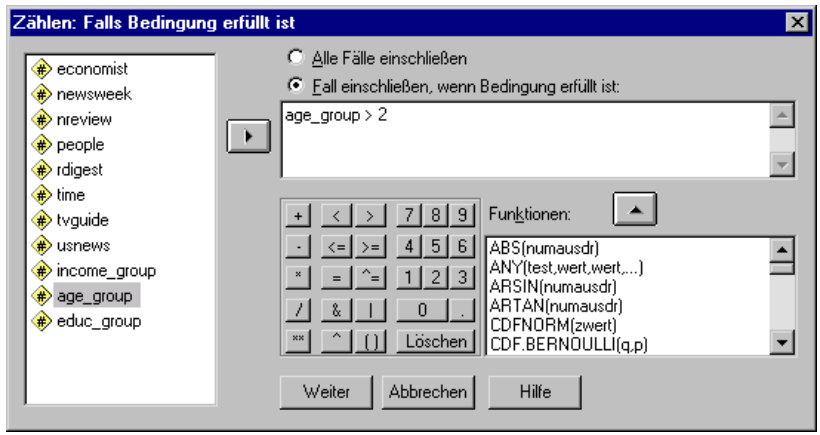
Abbildung 8-6
Dialogfeld "Zu zählende Werte"



Häufigkeiten von Werten in Fällen zählen: Falls Bedingung erfüllt ist

Im Dialogfeld "Variable berechnen: Falls Bedingung erfüllt ist" können Sie die Häufigkeiten von Werten für Teilmengen von Fällen zählen, die anhand bedingter Ausdrücke ausgewählt werden. Ein bedingter Ausdruck gibt für jeden Fall den Wert *Wahr*, *Falsch* oder *Fehlend* zurück.

Abbildung 8-7
Dialogfeld "Häufigkeiten von Werten in Fällen zählen: Falls Bedingung erfüllt ist"



Allgemeine Erläuterungen zur Verwendung eines Dialogfelds “Falls Bedingung erfüllt ist” finden Sie unter “Variable berechnen: Falls Bedingung erfüllt ist” auf S. 167.

Umkodieren von Werten

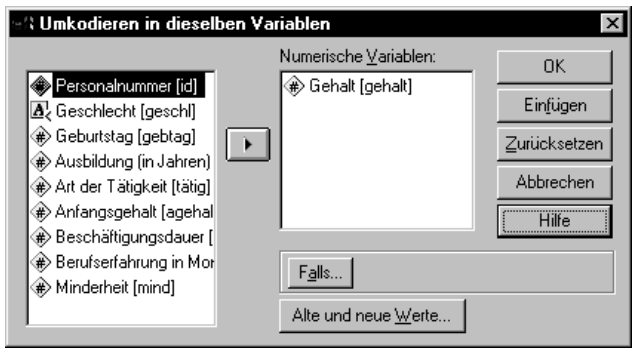
Sie können Datenwerte ändern, indem Sie diese umkodieren. Dies ist besonders nützlich, wenn Sie Kategorien zusammenfassen oder kombinieren. Sie können Werte in vorhandenen Variablen umkodieren oder neue Variablen auf der Grundlage der umkodierten Werte vorhandener Variablen erzeugen.

Umkodieren in dieselben Variablen

Im Dialogfeld “Umkodieren in dieselben Variablen” werden die Werte vorhandener Variablen erneut zugewiesen oder Bereiche vorhandener Werte in neuen Werten zusammengefaßt. So können Sie zum Beispiel Löhne in Kategorien von Lohnbereichen zusammenfassen.

Sie können numerische und String-Variablen umkodieren. Wenn Sie mehrere Variablen auswählen, müssen diese vom gleichen Typ sein. Sie können nicht numerische und String-Variablen gemeinsam umkodieren.

Abbildung 8-8
Dialogfeld “Umkodieren in dieselben Variablen”



So kodieren Sie die Werte einer Variablen um:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Transformieren
  Umkodieren
    In dieselben Variablen…

▶ Wählen Sie die Variablen aus, die Sie umkodieren möchten. Wenn Sie mehrere Variablen auswählen, müssen diese vom gleichen Typ (numerische oder String-Variablen) sein.

▶ Klicken Sie auf Alte und neue Werte, und geben Sie an, wie die Werte umkodiert werden sollen.

Wahlweise können Sie eine Teilmenge von Fällen zum Umkodieren definieren. Das Dialogfeld "Falls Bedingung erfüllt ist" für diese Funktion ist mit dem entsprechenden Dialogfeld für das Zählen der Häufigkeiten identisch.

Umkodieren in dieselben Variablen: Alte und neue Werte

In diesem Dialogfeld können Sie Werte zum Umkodieren definieren. Alle angegebenen Werte müssen vom gleichen Datentyp (numerisch oder String) wie die im Hauptdialogfeld ausgewählte Variable sein.

Alter Wert. Der (die) umzukodierende(n) Wert(e). Sie können einzelne Werte, Wertebereiche und fehlende Werte umkodieren. Für String-Variablen können keine systemdefinierten fehlenden Werte und Bereiche ausgewählt werden, da keines der beiden Konzepte auf String-Variablen zutrifft. Bei Bereichen sind die Endwerte und alle benutzerdefinierten fehlenden Werte eingeschlossen, die in den Bereich fallen.

- **Wert (Recode).** Einzelner alter Wert, der in einen neuen Wert umkodiert wird. Der Wert muß vom gleichen Datentyp sein (numerisch oder String) wie die umzukodierenden Variable(n).
- **Systemdefiniert fehlend.** Werte, die von SPSS zugewiesen werden, wenn Werte in den Daten nicht dem festgelegten Formattyp entsprechen, numerische Felder leer sind oder ein aus einem Transformationsbefehl resultierender Wert nicht definiert ist. Numerische systemdefinierte fehlende Werte werden durch Kommata dargestellt. String-Variablen können keine systemdefinierten fehlenden Werte enthalten, da in String-Variablen alle Zeichen zulässig sind.

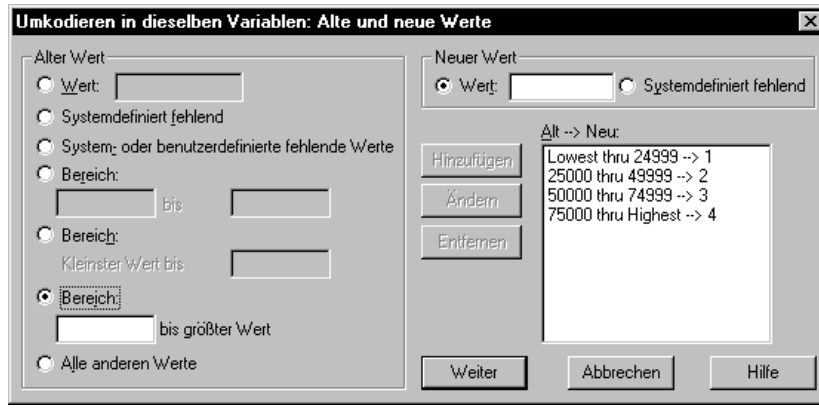
- **System- oder benutzerdefinierte fehlende Werte.** Beobachtungen mit Werten, die entweder als benutzerdefinierte fehlende Werte definiert wurden oder unbekannt sind und den systemdefinierten fehlenden Wert erhalten haben, der durch ein Komma (,) dargestellt wird.
- **Bereich (Inclusive Range of Values).** Einschließender Wertebereich. Nicht verfügbar für String-Variablen. Alle benutzerdefinierten, fehlenden Werte innerhalb des Bereichs sind eingeschlossen.
- **Alle anderen Werte.** Alle verbleibenden Werte, die nicht in den Angabe in der Alt-Neu-Liste enthalten sind. Dies erscheint als ELSE-Anweisung in der Alt-Neu-Liste.

Neuer Wert. Der einzelne Wert, in den jeder alte Wert oder Wertebereich umkodiert wird. Sie können einen Wert eingeben oder den systemdefiniert fehlenden Wert zuweisen.

- **Wert.** Der Wert, in den ein alter oder mehrere alte Werte umkodiert werden. Der neue Wert muß vom gleichen Datentyp (numerisch oder String) sein wie der alte Wert.
- **Systemdefiniert fehlend.** Kodiert die angegebenen alten Werte in den systemdefinierten fehlenden Wert um. Der systemdefinierte fehlende Wert wird bei Berechnungen nicht verwendet. Fälle mit dem systemdefinierten fehlenden Wert werden von vielen Prozeduren ausgeschlossen. Nicht verfügbar für String-Variablen.

Alt->Neu. Die Liste mit den Festlegungen, die zum Umkodieren von Variablen benötigt werden. Sie können Angaben hinzufügen, ändern und aus der Liste entfernen. Die Liste wird automatisch auf der Grundlage der alten Wertangaben sortiert. Dabei wird in folgender Reihenfolge vorgegangen: einzelne Werte, fehlende Werte, Bereiche und alle anderen Werte. Wenn Sie eine Angabe zum Umkodieren in der Liste ändern, sortiert SPSS die Liste gegebenenfalls automatisch neu, um diese Reihenfolge beizubehalten.

Abbildung 8-9
Dialogfeld "Alte und neue Werte"

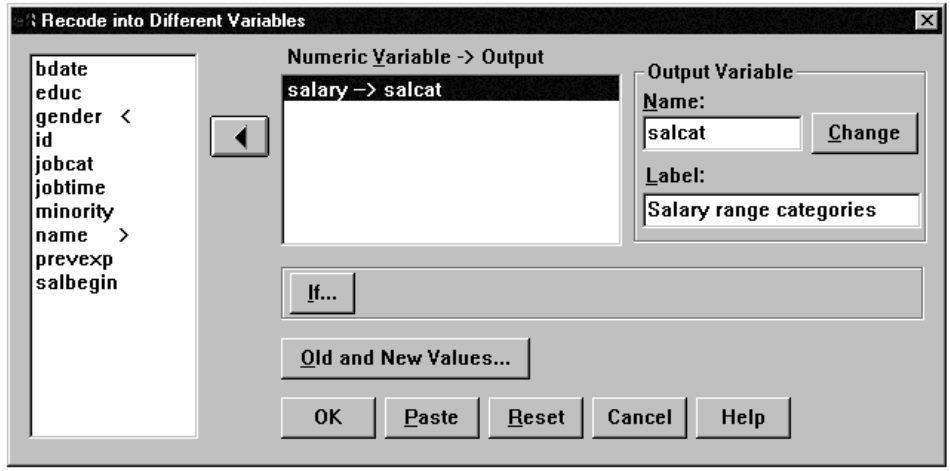


Umkodieren in andere Variablen

Im Dialogfeld "Umkodieren in andere Variablen" werden die Werte vorhandener Variablen erneut zugewiesen oder Bereiche vorhandener Werte in neuen Werten für eine neue Variable zusammengefaßt. So können Sie zum Beispiel Löhne zu einer neuen Variablen mit Lohnbereichen zusammenfassen.

- Sie können numerische und String-Variablen umkodieren.
- Sie können numerische Variablen in String-Variablen umkodieren und umgekehrt.
- Wenn Sie mehrere Variablen auswählen, müssen diese vom gleichen Typ sein. Sie können nicht numerische und String-Variablen gemeinsam umkodieren.

Abbildung 8-10
Dialogfeld "Umkodieren in andere Variablen"



So kodieren Sie die Werte einer Variablen in eine neue Variable um:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Transformieren
  Umkodieren
  In andere Variablen…

- Wählen Sie die Variablen aus, die Sie umkodieren möchten. Wenn Sie mehrere Variablen auswählen, müssen diese vom gleichen Typ (numerische oder String-Variablen) sein.

- Geben Sie für jede neue Variable einen neuen Namen an, und klicken Sie auf Ändern.

- Klicken Sie auf Alte und neue Werte, und geben Sie an, wie die Werte umkodiert werden sollen.

  Wahlweise können Sie eine Teilmenge von Fällen zum Umkodieren definieren. Das Dialogfeld "Falls Bedingung erfüllt ist" für diese Funktion ist mit dem entsprechenden Dialogfeld für das Zählen der Häufigkeiten identisch.

Umkodieren in andere Variablen: Alte und neue Werte

In diesem Dialogfeld können Sie Werte zum Umkodieren definieren.

Alter Wert. Der (die) umzukodierende(n) Wert(e). Sie können einzelne Werte, Wertebereiche und fehlende Werte umkodieren. Für String-Variablen können keine systemdefinierten fehlenden Werte und Bereiche ausgewählt werden, da keines der beiden Konzepte auf String-Variablen zutrifft. Die alten Werte müssen vom gleichen Datentyp (numerisch oder String) wie die ursprüngliche Variable sein. Bei Bereichen sind die Endwerte und alle benutzerdefinierten fehlenden Werte eingeschlossen, die in den Bereich fallen.

- **Wert (Recode).** Einzelner alter Wert, der in einen neuen Wert umkodiert wird. Der Wert muß vom gleichen Datentyp sein (numerisch oder String) wie die umzukodierenden Variable(n).
- **Systemdefiniert fehlend.** Werte, die von SPSS zugewiesen werden, wenn Werte in den Daten nicht dem festgelegten Formattyp entsprechen, numerische Felder leer sind oder ein aus einem Transformationsbefehl resultierender Wert nicht definiert ist. Numerische systemdefinierte fehlende Werte werden durch Kommata dargestellt. String-Variablen können keine systemdefinierten fehlenden Werte enthalten, da in String-Variablen alle Zeichen zulässig sind.
- **System- oder benutzerdefinierte fehlende Werte.** Beobachtungen mit Werten, die entweder als benutzerdefinierte fehlende Werte definiert wurden oder unbekannt sind und den systemdefinierten fehlenden Wert erhalten haben, der durch ein Komma (,) dargestellt wird.
- **Bereich (Inclusive Range of Values).** Einschließender Wertebereich. Nicht verfügbar für String-Variablen. Alle benutzerdefinierten, fehlenden Werte innerhalb des Bereichs sind eingeschlossen.
- **Alle anderen Werte.** Alle verbleibenden Werte, die nicht in den Angabe in der Alt-Neu-Liste enthalten sind. Dies erscheint als ELSE-Anweisung in der Alt-Neu-Liste.

Neuer Wert. Der einzelne Wert, in den jeder alte Wert oder Wertebereich umkodiert wird. Die neuen Werte können numerische oder String-Variablen sein.

- **Wert.** Der Wert, in den ein alter oder mehrere alte Werte umkodiert werden. Der neue Wert muß vom gleichen Datentyp (numerisch oder String) sein wie der alte Wert.
- **Systemdefiniert fehlend.** Kodiert die angegebenen alten Werte in den systemdefinierten fehlenden Wert um. Der systemdefinierte fehlende Wert wird bei Berechnungen nicht verwendet. Fälle mit dem systemdefinierten fehlenden

Wert werden von vielen Prozeduren ausgeschlossen. Nicht verfügbar für String-Variablen.

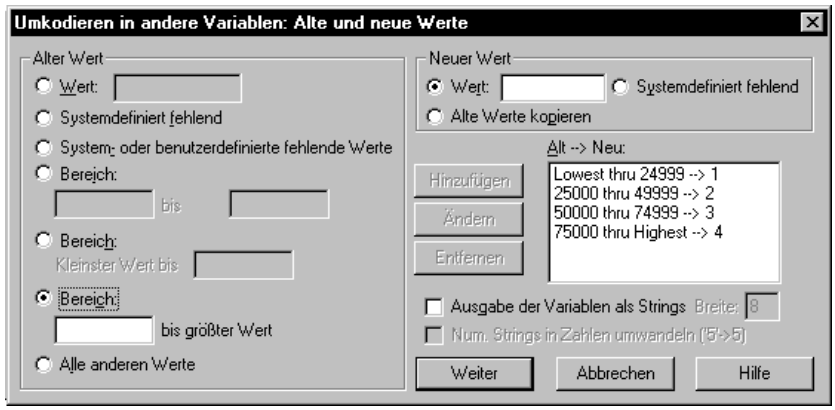
- **Kopieren alter Werte.** Erhält die alten Werte. Falls einige Werte keine Umkodierung benötigen, verwenden Sie diese Funktion um alte Werte einzuschließen. Alle alten Werte, die nicht festgelegt sind, werden nicht in den neuen Variablen eingeschlossen, und Fällen mit diesen Werten wird der systemdefinierte fehlende Wert für die neue Variable zugewiesen.

Ausgabe der Variablen als Strings. Definiert die neue, umkodierte Variable als String-Variable (alphanumerische Variable). Die alte Variable kann eine numerische Variable oder eine String-Variable sein.

Umwandeln numerischer Strings in Zahlen. Konvertiert String-Werte, die Zahlen enthalten, in numerische Werte. Strings, die andere Zeichen als Zahlen und ein optionales Vorzeichen (+ oder -) enthalten, wird der systemdefinierte fehlende Wert zugewiesen.

Alt->Neu. Die Liste mit den Festlegungen, die zum Umkodieren von Variablen benötigt werden. Sie können Angaben hinzufügen, ändern und aus der Liste entfernen. Die Liste wird automatisch auf der Grundlage der alten Wertangaben sortiert. Dabei wird in folgender Reihenfolge vorgegangen: einzelne Werte, fehlende Werte, Bereiche und alle anderen Werte. Wenn Sie eine Angabe zum Umkodieren in der Liste ändern, sortiert SPSS die Liste gegebenenfalls automatisch neu, um diese Reihenfolge beizubehalten.

Abbildung 8-11
Dialogfeld "Alte und neue Werte"



Rangfolge bilden

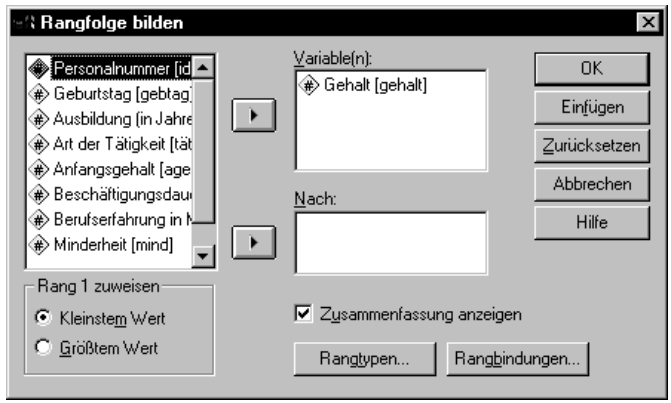
Im Dialogfeld “Fällen Rang zuordnen” werden für numerische Variablen neue Variablen mit Rängen, Normalrangwerten, Savage-Werten und Perzentil-Werten erstellt.

Neue Variablennamen und beschreibende Variablenlabels werden von SPSS automatisch auf der Grundlage des ursprünglichen Variablennamens und der ausgewählten Maße erzeugt. In einer Übersichtstabelle werden die ursprünglichen Variablen, die neuen Variablen und die Variablenlabels aufgelistet.

Die folgenden Optionen sind verfügbar:

- Rangfolge der Fälle in aufsteigender oder absteigender Reihenfolge bilden.
- Rangfolgen in Untergruppen organisieren, indem Sie mindestens eine Gruppenvariable für die Liste “Nach” auswählen. Ränge werden innerhalb jeder Gruppe berechnet. Gruppen werden durch die Kombination der Werte der Gruppenvariablen definiert. Wenn Sie zum Beispiel *Geschlecht* und *Minderheit* als Gruppenvariablen auswählen, werden Ränge für jede Kombination von *Geschlecht* und *Minderheit* berechnet.

Abbildung 8-12
Dialogfeld “Fällen Rang zuordnen”



So bilden Sie die Rangfolge der Fälle

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Transformieren
  Rangfolge bilden…

- Wählen Sie mindestens eine Variable, für die Sie eine Rangfolge bilden möchten. Sie können nur Rangfolgen von numerischen Variablen bilden.

  Sie können die Ränge der Fälle wahlweise in aufsteigender oder absteigender Reihenfolge anordnen und Ränge in Untergruppen organisieren.

Rangfolge bilden: Typen

Zum Bilden der Rangfolge stehen mehrere Methoden zur Verfügung. Für jede Methode wird eine separate Rangvariable erstellt. Beim Bilden der Rangfolge können einfache Ränge, Savage-Werte, relative Ränge und Perzentile eingesetzt werden. Sie können die Rangfolge auch auf der Grundlage von Anteilsschätzungen und Normalrangwerten erzeugen.

Rang. Einfacher Rang. Der Wert der neuen Variablen ist gleich ihrem Rang.

Savage-Wert. Die neue Variable enthält Savage-Werte auf der Grundlage einer exponentiellen Verteilung.

Relative Rangfolge. Der Wert der neuen Variable ist gleich dem Rangwert geteilt durch die Summe der Gewichtungen nichtfehlender Fälle.

Relative Rangfolge in Prozent. Jeder Rangwert wird geteilt durch die Anzahl der Fälle mit gültigen Werten und multipliziert mit 100.

Summe der Fallgewichtungen. Der Wert der neuen Variablen ist gleich der Summe der Fallgewichte. Die neue Variable ist für alle Fälle in derselben Gruppe eine Konstante.

N-Perzentile. Die Ränge basieren auf Perzentilgruppen, wobei jede Gruppe ungefähr die gleiche Anzahl von Fällen enthält. So erhalten beispielsweise bei 4 N-Perzentilen Fälle unter dem 25. Perzentil den Rang 1, Fälle zwischen dem 25. und 50. Perzentil den Rang 2, Fälle zwischen dem 50. und 75. Perzentil den Rang 3 und Fälle über dem 75. Perzentil den Rang 4.

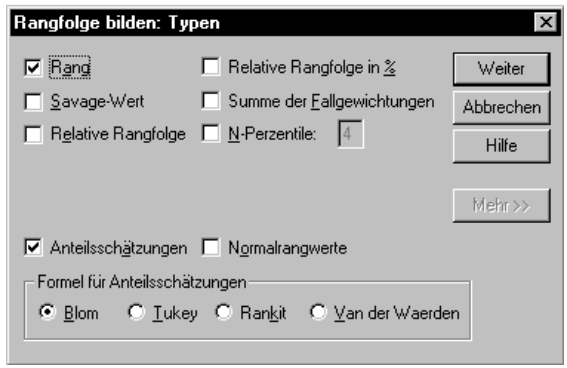
Anteilsschätzungen. Anteilsschätzer sind Schätzungen des kumulierten Anteils der Verteilung bezüglich eines einzelnen Ranges.

Normalrangwerte. Die Z-Werte, welche dem geschätzten kumulativen Anteil entsprechen.

Formel für Anteilsschätzungen. Für Anteilsschätzungen und Normalrangwerte können Sie die Formel für die Anteilsschätzung auswählen: Blom, Tukey, Rankit oder Van der Waerden.

- **Blom.** Erstellt eine neue Rangvariable auf der Grundlage von mit der Formel (r-3/8) / (w+1/4) berechneten Anteilsschätzern, wobei w die Summe der Fallgewichtungen und r der Rang ist.
- **Tukey.** Verwendet die Formel (r-1/3) / (w+1/3), wobei r den Rang und w die Summe der Fallgewichtungen angibt.
- **Rankit.** Es wird die Formel (r-1/2) / w verwendet, wobei w die Anzahl der Beobachtungen und r der Rang ist, der von 1 bis w reicht.
- **Van der Waerden.** Durch die Formel r/(w+1) definierte Van der Waerden-Transformation, wobei w die Summe der Fallgewichte und r den von 1 bis n reichenden Rang darstellt.

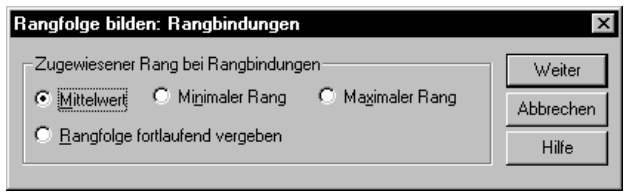
Abbildung 8-13
Dialogfeld "Rangfolge bilden: Typen"



Rangfolge bilden: Bindungen

In diesem Dialogfeld werden Einstellungen für die Methode zum Zuweisen von Rängen zu Fällen mit demselben Wert in der ursprünglichen Variablen vorgenommen.

Abbildung 8-14
Dialogfeld "Fällen Rang zuordnen: Rangbindungen"



Die folgende Tabelle zeigt, wie den gebundenen Werten bei verschiedenen Methoden Ränge zugewiesen werden.

Wert	Mittelwert	Kleinster Wert	Größter Wert	Fortlaufend
10	1	1	1	1
15	3	2	4	2
15	3	2	4	2
15	3	2	4	2
16	5	5	5	3
20	6	6	6	4

Automatisch umkodieren

Im Dialogfeld “Automatisch umkodieren” wandeln Sie String-Werte und numerische Werte in fortlaufende Ganzzahlen um. Wenn Kategoriecodes nicht sequentiell sind, vermindern die daraus resultierenden leeren Zellen die Leistung und erhöhen den Speicherbedarf für viele SPSS-Prozeduren. Außerdem können einige Prozeduren keine String-Variablen verwenden, und einige erfordern aufeinanderfolgende ganzzahlige Werte als Faktorstufen.

Abbildung 8-15
Dialogfeld "Automatisch umkodieren"

- Bei den mit "Automatisch umkodieren" erstellten neuen Variablen werden alle definierten Variablen und Wertelabels aus der alten Variablen beibehalten. Bei allen Werten ohne definiertes Wertelabel wird der ursprüngliche Wert als Label für den umkodierten Wert verwendet. Die alten und neuen Werte und Wertelabels werden in einer Tabelle angezeigt.
- String-Werte werden in alphabetischer Reihenfolge umkodiert, wobei Großbuchstaben jeweils vor den entsprechenden Kleinbuchstaben stehen.
- Fehlende Werte werden unter Beibehaltung ihrer Reihenfolge in fehlende Werte umkodiert, die größer als alle nichtfehlenden Werte sind. Wenn zum Beispiel die ursprüngliche Variable über 10 nichtfehlende Werte verfügt, würde der kleinste fehlende Wert auf 11 umkodiert, und der Wert 11 wäre ein fehlender Wert für die neue Variable.

Dasselbe Umkodierungsschema für alle Variablen verwenden. Mit dieser Option können Sie ein einziges Schema für die automatische Umkodierung auf alle ausgewählten Variablen anwenden und so ein einheitliches Kodierungsschema für alle neuen Variablen erzielen.

Bei dieser Option gelten die folgenden Regeln und Einschränkungen:

- Alle Variablen müssen denselben Typ aufweisen (numerische Variable oder String-Variable).
- Alle beobachteten Werte für alle ausgewählten Variablen dienen als Grundlage für eine Sortierreihenfolge, mit der die Werte vor der Umkodierung in fortlaufende Ganzzahlen sortiert werden.
- Benutzerdefiniert fehlende Werte für die neuen Variablen beruhen auf der ersten Variable in der Liste mit angegebenen benutzerdefiniert fehlenden Werten. Alle anderen Werte aus anderen ursprünglichen Variablen (mit Ausnahme der systemdefiniert fehlenden Werte) werden als gültige Werte behandelt.

Leerstring-Werte als benutzerdefiniert fehlend behandeln. Bei String-Variablen gelten leere Werte oder Null-Werte nicht als **systemdefiniert fehlend**. Mit dieser Option werden leere Strings automatisch in einen **benutzerdefinierten fehlenden** Wert größer als der größte nichtfehlende Wert umkodiert.

Vorlagen zum Definieren von Variablen

Sie können das Schema für die automatische Kodierung in einer Vorlagendatei speichern und dann auf andere Variablen und andere Datendateien anwenden.

Sie verwenden beispielsweise zahlreiche alphanumerische Produktcodes, die Sie jeden Monat automatisch in Ganzzahlen umkodieren lassen. In einigen Monaten werden jedoch neue Produktcodes eingeführt, die das ursprüngliche Schema für die automatische Umkodierung ändern. Wenn Sie das ursprüngliche Schema in einer Vorlage speichern und dann auf die neuen Daten anwenden, die die neuen Codes enthalten, werden alle neuen Codes in den Daten automatisch in Werte umkodiert, die höher sind als der höchste Wert in der Vorlage. Auf diese Weise wird das ursprüngliche Schema für die automatische Umkodierung der ursprünglichen Produktcodes beibehalten.

Vorlage speichern als. Speichert das Schema für die automatische Umkodierung der ausgewählten Variablen in einer externen Vorlagendatei.

- Mit den Informationen in der Vorlage werden die ursprünglichen nichtfehlenden Werte den umkodierten Werten zugeordnet.
- Nur Informationen für nichtfehlende Werte werden in der Vorlage gespeichert. Informationen zu benutzerdefiniert fehlenden Werten werden nicht beibehalten.

- Wenn Sie mehrere Variablen für die Umkodierung ausgewählt haben, ohne dabei dasselbe Schema für die automatische Umkodierung für alle Variablen festzulegen oder wenn keine vorhandene Vorlage im Rahmen der automatischen anzuwenden, wird die Vorlage auf der Grundlage der ersten Variable in der Liste aufgebaut.
- Wenn Sie mehrere Variablen für die Umkodierung ausgewählt und dabei die Option Dasselbe Umkodierungsschema für alle Variablen verwenden und/oder die Option Vorlage zuweisen aktiviert haben, enthält die Vorlage das kombinierte Schema für die automatische Umkodierung für alle Variablen.

Vorlage übernehmen aus. Wendet eine zuvor gespeicherte Vorlage für die automatische Umkodierung auf alle Variablen an, die zur Umkodierung ausgewählt wurden. Alle zusätzlich in den Variablen gefundenen Werte werden an das Ende des Schemas angehängt, so daß die Beziehung zwischen den ursprünglichen und den automatisch umkodierten Werten erhalten bleibt, die im gespeicherten Schema festgehalten ist.

- Alle zur Umkodierung ausgewählten Variablen müssen denselben Typ aufweisen (numerische Variable oder String-Variable), und dieser Typ muß mit dem Typ übereinstimmen, der in der Vorlage definiert ist.
- Vorlagen enthalten keine Informationen zu benutzerdefiniert fehlenden Werten. Benutzerdefiniert fehlende Werte für die Zielvariablen beruhen auf der ersten Variable in der Liste mit angegebenen benutzerdefiniert fehlenden Werten. Alle anderen Werte aus anderen ursprünglichen Variablen (mit Ausnahme der systemdefiniert fehlenden Werte) werden als gültige Werte behandelt.
- Wertzuordnungen aus der Vorlage werden als erstes angewendet. Alle verbleibenden Werte werden in Werte umkodiert, die höher sind als der letzte Wert in der Vorlage. Benutzerdefiniert fehlende Werte (auf der Grundlage der ersten Variable in der Liste mit definierten benutzerdefiniert fehlenden Werten) werden dabei in Werte umkodiert, die höher sind als der letzte gültige Wert.
- Wenn Sie mehrere Variablen für die automatische Umkodierung ausgewählt haben, wird zunächst die Vorlage angewendet. Anschließend wird eine kombinierte automatische Standard-Umkodierung für alle zusätzlichen Werte für die ausgewählten Variablen ausgeführt. So entsteht ein einziges gemeinsames Schema für die automatische Umkodierung, das alle ausgewählten Variablen erfaßt.

So kodieren Sie String- oder numerische Werte in fortlaufende Ganzzahlen um

- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Transformieren
  Automatisch umkodieren
- Wählen Sie mindestens eine Variable zum Umkodieren aus.
- Geben Sie für jede ausgewählte Variable einen Namen für die neue Variable ein, und klicken Sie auf Neuer Name.

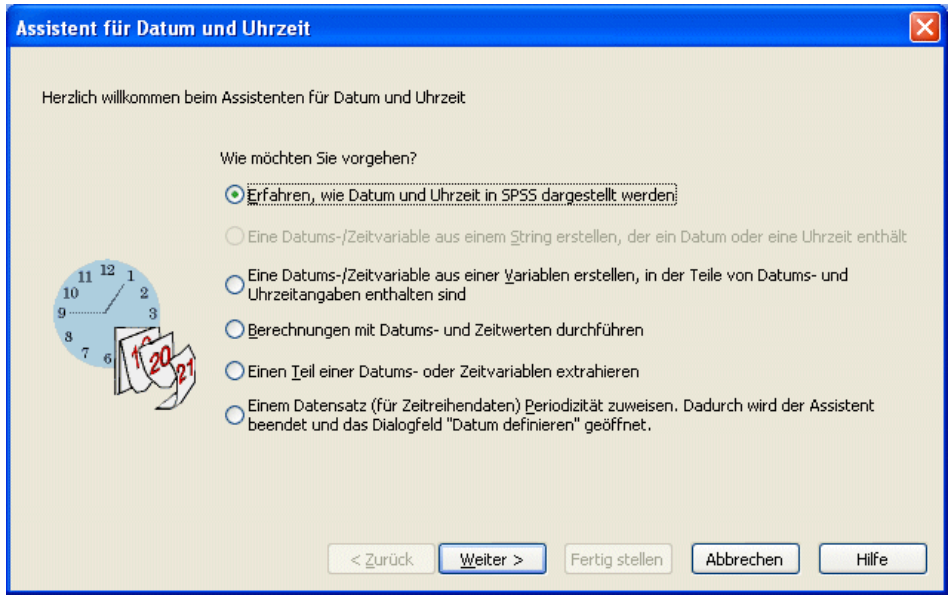
Assistent für Datum und Uhrzeit

Der Assistent für Datum und Uhrzeit vereinfacht eine Reihe von Aufgaben im Zusammenhang mit Datums- und Zeitvariablen.

So verwenden Sie den Assistenten für Datum und Uhrzeit

- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Transformieren
  Datum/Uhrzeit...
- Wählen Sie die gewünschte Aufgabe aus, und befolgen Sie die Schritte zur Definition der Aufgabe.

Abbildung 8-16
Assistent für Datum und Uhrzeit: Einführungsbildschirm


Assistent für Datum und Uhrzeit
Herzlich willkommen beim Assistenten für Datum und Uhrzeit
Wie möchten Sie vorgehen?
Erfahren, wie Datum und Uhrzeit in SPSS dargestellt werden
Eine Datums-/Zeitvariable aus einem String erstellen, der ein Datum oder eine Uhrzeit enthält
Eine Datums-/Zeitvariable aus einer Variablen erstellen, in der Teile von Datums- und Uhrzeitangaben enthalten sind
Berechnungen mit Datums- und Zeitwerten durchführen
Einen Teil einer Datums- oder Zeitvariablen extrahieren
Einem Datensatz (für Zeitreihendaten) Periodizität zuweisen. Dadurch wird der Assistent beendet und das Dialogfeld "Datum definieren" geöffnet.
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

- **Erfahren, wie Datum und Uhrzeit in SPSS dargestellt werden.** Diese Option führt Sie zu einem Bildschirm, in dem ein kurzer Überblick über die Datums-/Zeitvariablen in SPSS geboten wird. Wenn Sie auf Hilfe klicken, wird außerdem eine Verknüpfung zu detaillierteren Informationen angezeigt.
- **Eine Datums-/Zeitvariable aus einem String erstellen, der ein Datum oder eine Uhrzeit enthält.** Mit dieser Option können Sie eine Datums-/Zeitvariable aus einer String-Variablen erstellen. Beispiel: Sie haben eine String-Variable mit Datumsangaben im Format mm/tt/jjjj und möchten daraus eine Datums-/Zeitvariable erstellen.
- **Eine Datums-/Zeitvariable aus einer Variablen erstellen, in der Teile von Datums- und Uhrzeitangaben enthalten sind.** Mit dieser Option können Sie eine Datums-/Zeitvariable aus einem Set bestehender Variablen erstellen. Beispiel: Sie haben eine Variable für den Monat (als Ganzzahl), eine weitere Variable für den Tag im Monat und eine dritte für das Jahr. Sie können diese drei Variablen zu einer einzigen Datums-/Zeitvariablen kombinieren.

- **Berechnungen mit Datums- und Zeitwerten durchführen.** Diese Option dient zum Addieren oder Subtrahieren von Werten zu bzw. von Datums-/Zeitvariablen. Beispiel: Sie können die Dauer eines Prozesses berechnen, indem Sie eine Variable, die die Startzeit des Prozesses angibt, von einer anderen Variablen subtrahieren, die den Zeitpunkt des Prozeßendes angibt.
- **Einen Teil einer Datums- oder Zeitvariablen extrahieren.** Mit dieser Option können Sie einen Teil einer Datums-/Zeitvariablen extrahieren, beispielsweise des Tages im Monat aus einer Datums-/Zeitvariablen mit dem Format mm/tt/jjjj.
- **Einem Datensatz Periodizität zuweisen.** Mit dieser Option gelangen Sie zum Dialogfeld "Datum definieren", das zum Erstellen von Datums-/Zeitvariablen verwendet wird, die aus einer Reihe aufeinanderfolgender Datumsangaben bestehen. Diese Funktion wird normalerweise verwendet, um Datumsangaben Zeitreihendaten zuzuweisen.

Anmerkung: Aufgaben werden deaktiviert, wenn das Daten-Set nicht die für die Ausführung der Aufgabe erforderlichen Variablen aufweist. Wenn das Daten-Set beispielsweise keine String-Variablen enthält, findet die Aufgabe zur Erstellung einer Datums-/Zeitvariablen aus einer String-Variablen keine Anwendung und ist deaktiviert.

Datums- und Zeitangaben in SPSS

Variablen für Datums- und Zeitangaben in SPSS weisen einen numerischen Variablentyp auf, mit Anzeigeformaten die den jeweiligen Datums-/Zeitformaten entsprechen. Diese Variablen werden im allgemeinen als Datums-/Zeitvariablen bezeichnet. In SPSS wird zwischen Datums-/Zeitvariablen unterschieden, die tatsächlich für einen bestimmten Datumswert stehen, und solchen, die eine Zeitdauer repräsentieren, die unabhängig von einem bestimmten Datum ist, wie beispielsweise 20 Stunden, 10 Minuten und 15 Sekunden. Letztere werden als Dauer-Variablen und erster als Datums- oder Datums-/Zeitvariablen bezeichnet. Eine Vollständige Liste der Anzeigeformate finden Sie in der *SPSS Command Syntax Reference*, im Abschnitt "Universals" unter "Date and Time".

Datums- und Datums-/Zeitvariablen. Datumsvariablen weisen ein Format auf, das einem Datum entspricht, beispielsweise mm/tt/jjjj. Datums-/Zeitvariablen weisen ein Format auf, das einem Datum und einer Uhrzeit entspricht, beispielsweise tt-mmm-jjjj hh:mm:ss. Intern werden Datums- und Datums-/Zeitvariablen als die Anzahl der

seit dem 14. Oktober 1582 vergangenen Sekunden gespeichert. Datums- und Datums-/Zeitvariablen werden manchmal als Variablen mit Datumsformat bezeichnet.

- Jahresangaben werden in zweistelligen und im vierstelligen Format erkannt. In der Standardeinstellung wird bei zweistelligen Jahreszahlen ein Bereich angenommen, der 69 Jahre vor dem gegenwärtigen Datum und 30 Jahre danach umfaßt. Dieser Bereich hängt von den Optionseinstellungen ab und kann konfiguriert werden (wählen Sie dazu im Menü "Bearbeiten" den Befehl Optionen aus, und klicken Sie auf die Registerkarte Daten).
- In Formaten vom Typ Tag-Monat-Jahr können Bindestriche, Punkte, Kommata, Schrägstriche und Leerzeichen als Trennzeichen verwendet werden.
- Monate können durch arabische oder römische Ziffern und aus drei Buchstaben bestehende Abkürzungen dargestellt oder vollständig ausgeschrieben werden. Abkürzungen aus drei Buchstaben und vollständig ausgeschriebene Monatsnamen müssen in englischer Sprache vorliegen; Monatsnamen in anderen Sprachen werden nicht erkannt.

Dauer-Variablen. Dauer-Variablen weisen ein Format auf, das einer Zeitdauer entspricht, beispielsweise hh:mm. Sie werden intern als Sekunden ohne Bezug auf ein bestimmtes Datum gespeichert.

- Bei Zeitangaben (gilt für Datums-/Zeit- und Dauer-Variablen) können Doppelpunkte als Trennzeichen zwischen Stunden, Minuten und Sekunden verwendet werden. Stunden und Minuten sind erforderlich, Sekunden dagegen sind optional. Ein Punkt ist erforderlich, um Sekunden von Sekundenbruchteilen zu trennen. Für Stunden kann ein beliebig hoher Wert angegeben werden. Der maximale Wert für die Minuten ist jedoch 59 und für die Sekunden 59.999...

Aktuelles Datum und aktuelle Uhrzeit. Die Systemvariable *$TIME* enthält das aktuelle Datum und die aktuelle Uhrzeit. Sie steht für die Anzahl der Sekunden seit dem 14. Oktober 1582 bis zu dem Datum und der Uhrzeit, zu der der Transformationsbefehl, der diese Variable verwendet, ausgeführt wird.

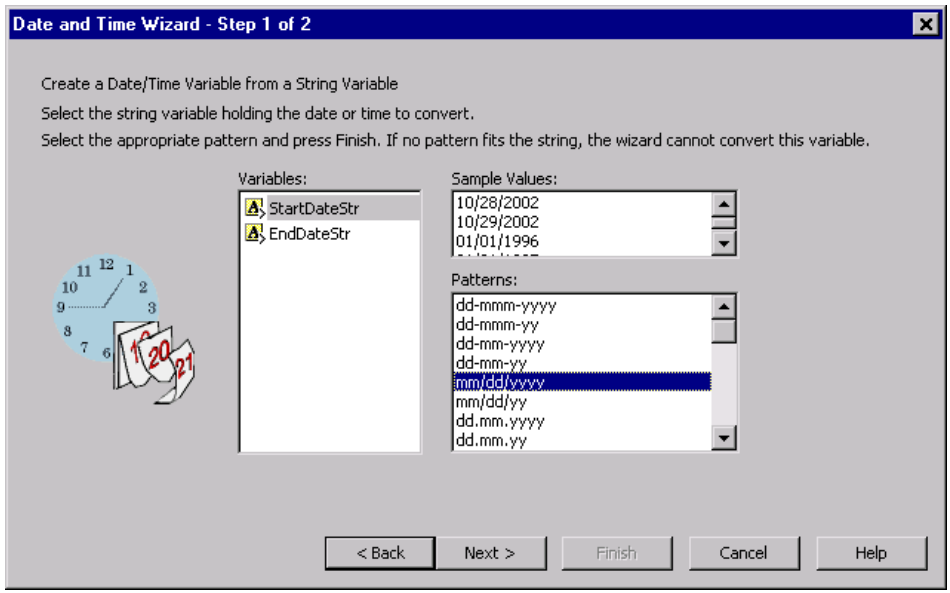
Erstellen einer Datums-/Zeitvariablen aus einer String-Variablen

So erstellen Sie eine Datums-/Zeitvariable aus einer String-Variablen:

- Wählen Sie im Einführungsbildschirm des Assistenten für Datum und Uhrzeit die Option Eine Datums-/Zeitvariable aus einem String erstellen, der ein Datum oder eine Uhrzeit enthält aus.

Auswählen einer String-Variablen zur Umwandlung in eine Datums-/Zeitvariable

Abbildung 8-17
Erstellen einer Datums-/Zeitvariablen aus einer String-Variablen, Schritt 1

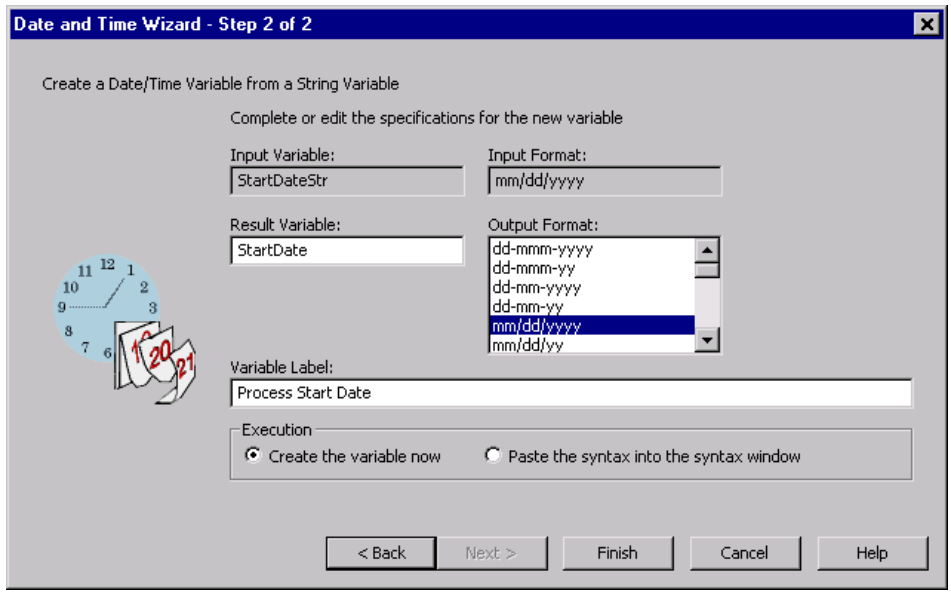


- Wählen Sie die umzuwandelnde String-Variable in der Variablenliste aus. Beachten Sie, daß in der Liste nur String-Variablen angezeigt werden.

- Wählen Sie aus der Liste “Muster” das Muster aus, das mit der Form der Datumsangaben in der String-Variablen übereinstimmt. In der Liste “Beispielwerte” werden die tatsächlichen Werte der ausgewählten Variablen in der Datendatei angezeigt. Werte der Stringvariablen, die nicht zum ausgewählten Muster passen, führen zum Wert “systemdefiniert fehlend” für die neue Variable.

Angeben des Ergebnisses der Umwandlung einer String-Variablen in eine Datums-/Zeitvariable

Abbildung 8-18
Erstellen einer Datums-/Zeitvariablen aus einer String-Variablen, Schritt 2



- Geben Sie einen Namen für die Ergebnisvariable ein. Dieser darf nicht mit dem Namen einer bestehenden Variablen übereinstimmen.

  Die folgenden Optionen sind verfügbar:

  - Auswählen eines Datums-/Uhrzeitformats für die neue Variable in der Liste “Ausgabeformat”
  - Zuweisen eines beschreibenden Variablenlabels zur neuen Variablen.

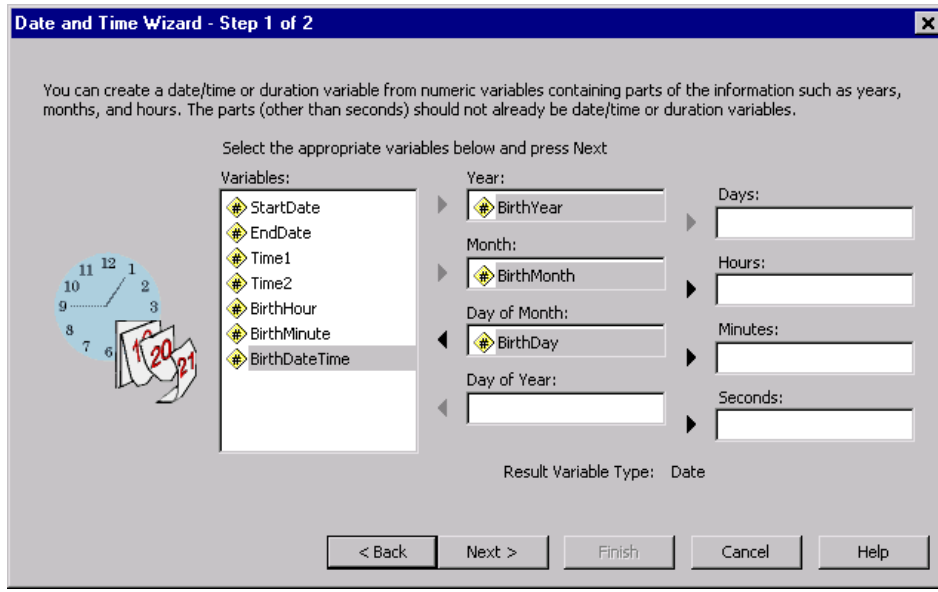
Erstellen einer Datums-/Zeitvariablen aus einem Variablen-Set

So führen Sie eine Menge bestehender Variablen zu einer einzigen Datums-/Zeitvariablen zusammen:

- Wählen Sie im Einführungsbildschirm des Assistenten für Datum und Uhrzeit die Option Eine Datums-/Zeitvariable aus einer Variablen erstellen, in der Teile von Datums- und Uhrzeitangaben enthalten sind aus.

Auswählen von Variablen zur Zusammenführung in eine einzige Datums-/Zeitvariable

Abbildung 8-19
Erstellen einer Datums-/Zeitvariablen aus einem Variablen-Set, Schritt 1

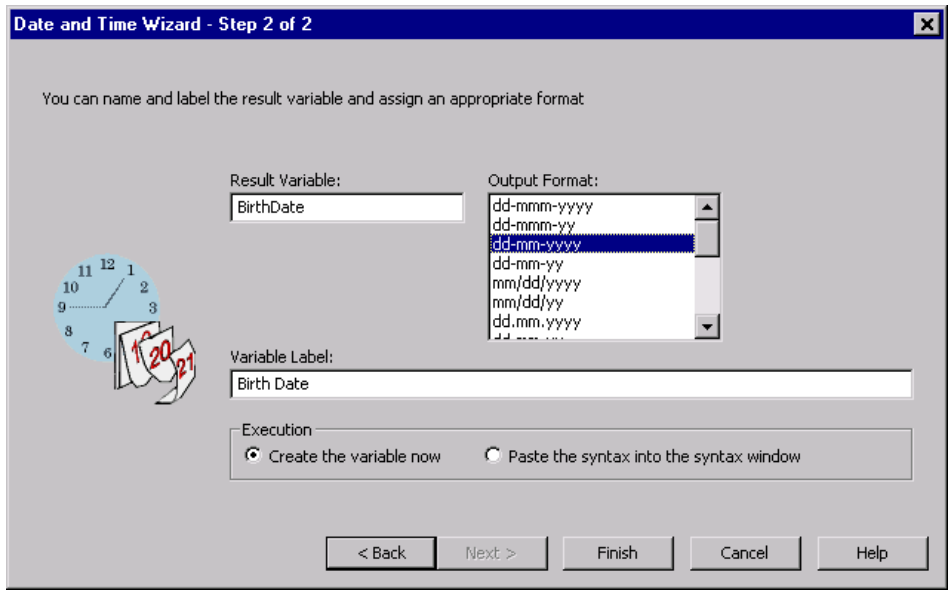


- Wählen Sie die Variablen aus, die für die verschiedenen Teile des Datums bzw. der Uhrzeit stehen.
  - Einige Kombinationen sind nicht zulässig. So kann beispielsweise keine Datums-/Zeitvariable aus “Jahr” und “Tag im Monat” erstellt werden, da nach der Auswahl von “Jahr” ein vollständiges Datum erforderlich ist.

- Bereits bestehende Datums-/Zeitvariablen können nicht als Teil der zu erstellenden Datums-/Zeitvariablen verwendet werden. Bei den Variablen für die Teile der neuen Datums-/Zeitvariablen muß es sich um Ganzzahlen handeln. Eine Ausnahme ist die zulässige Verwendung einer bestehenden Datums-/Zeitvariablen als Sekunden-Teil der neuen Variablen. Da Sekundenbruchteile zulässig sind, muß die für die Sekunden verwendete Variable keine Ganzzahl sein.
- Werte für einen Teil der neuen Variablen, die nicht innerhalb des zulässigen Bereichs liegen, führen zum Wert "systemdefiniert fehlend" für die neue Variable. Beispiel: Sie verwenden für den Monat versehentlich eine Variable, die den Tag im Monat angibt. Da der gültige Bereich für Monate in SPSS 1–13 beträgt, wird allen Fällen, bei denen der Wert für den Tag im Monat im Bereich von 14–31 liegt, der Wert "systemdefiniert fehelend" für die neue Variable zugewiesen.

Angeben der durch das Zusammenführen von Variablen erstellten Datums-/Zeitvariablen.

Abbildung 8-20
Erstellen einer Datums-/Zeitvariablen aus einem Variablen-Set, Schritt 2



▶ Geben Sie einen Namen für die Ergebnisvariable ein. Dieser darf nicht mit dem Namen einer bestehenden Variablen übereinstimmen.

- Wählen Sie ein Datums-/Uhrzeitformat aus der Liste “Ausgabeformat” aus.

  Die folgenden Optionen sind verfügbar:

  - Zuweisen eines beschreibenden Variablenlabels zur neuen Variablen.

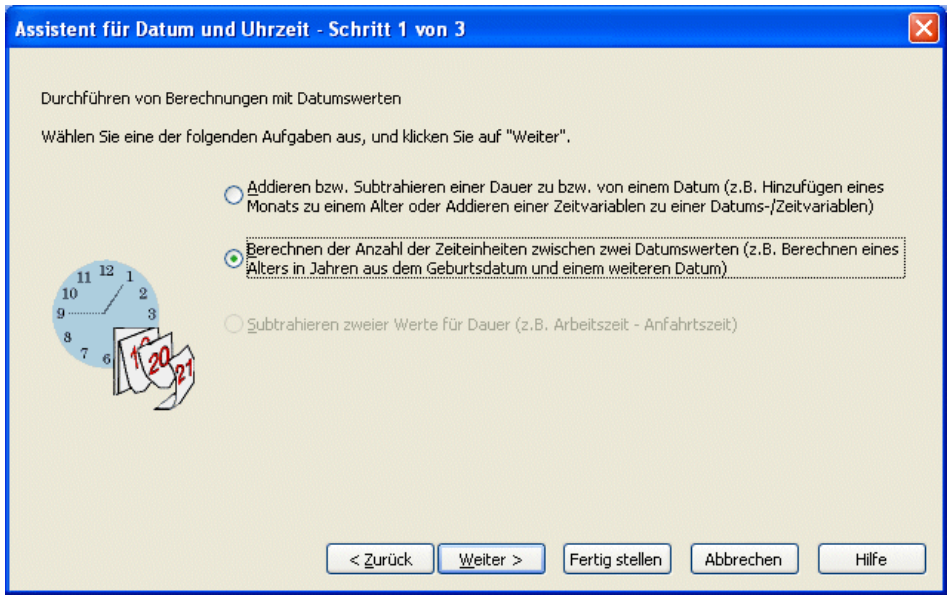
Addieren oder Subtrahieren von Werten zu bzw. von Datums-/Zeitvariablen

So können Sie Werte zu bzw. von Datums-/Zeitvariablen addieren bzw. subtrahieren:

- Wählen Sie im Einführungsbildschirm des Assistenten für Datum und Uhrzeit die Option Berechnungen mit Datums- und Zeitwerten durchführen aus.

Auswählen des Typs der mit Datums-/Zeitvariablen durchzuführenden Berechnung

Abbildung 8-21
Addieren oder Subtrahieren von Werten zu bzw. von Datums-/Zeitvariablen, Schritt 1


Assistent für Datum und Uhrzeit - Schritt 1 von 3
Durchführen von Berechnungen mit Datumswerten
Wählen Sie eine der folgenden Aufgaben aus, und klicken Sie auf "Weiter".
Addieren bzw. Subtrahieren einer Dauer zu bzw. von einem Datum (z.B. Hinzufügen eines Monats zu einem Alter oder Addieren einer Zeitvariablen zu einer Datums-/Zeitvariablen)
Berechnen der Anzahl der Zeiteinheiten zwischen zwei Datumswerten (z.B. Berechnen eines Alters in Jahren aus dem Geburtsdatum und einem weiteren Datum)
Subtrahieren zweier Werte für Dauer (z.B. Arbeitszeit - Anfahrtszeit)
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

- **Addieren bzw. Subtrahieren einer Dauer zu bzw. von einem Datum.** Diese Option dient zum Addieren oder Subtrahieren von Werten zu bzw. von Variablen mit Datumsformat. Sie können eine Zeitdauer, die einen festen Wert aufweist (z. B. 10 Tage), oder die Werte aus einer numerischen Variablen (beispielsweise einer Variablen, die Jahre angibt) addieren bzw. subtrahieren.

- **Berechnen der Anzahl der Zeiteinheiten zwischen zwei Datumswerten.** Mit dieser Option erhalten Sie die Differenz zwischen zwei Datumswerten (angegeben in einer von Ihnen ausgewählten Einheit). Beispielsweise können Sie die Anzahl der Jahre oder der Tage ermitteln, die zwischen zwei Datumsangaben liegt.
- **Subtrahieren zweier Werte für Dauer.** Mit dieser Option erhalten Sie die Differenz zwischen zwei Variablen, die ein Format für die Zeitdauer aufweisen, beispielsweise hh:mm oder hh:mm:ss.

Anmerkung: Aufgaben werden deaktiviert, wenn das Daten-Set nicht die für die Ausführung der Aufgabe erforderlichen Variablen aufweist. Beispiel: Wenn das Daten-Set nicht zwei Variablen mit einem Format für Zeitdauern aufweist, findet die Aufgabe zur Subtraktion zweier Werte für Dauer keine Anwendung und ist deaktiviert.

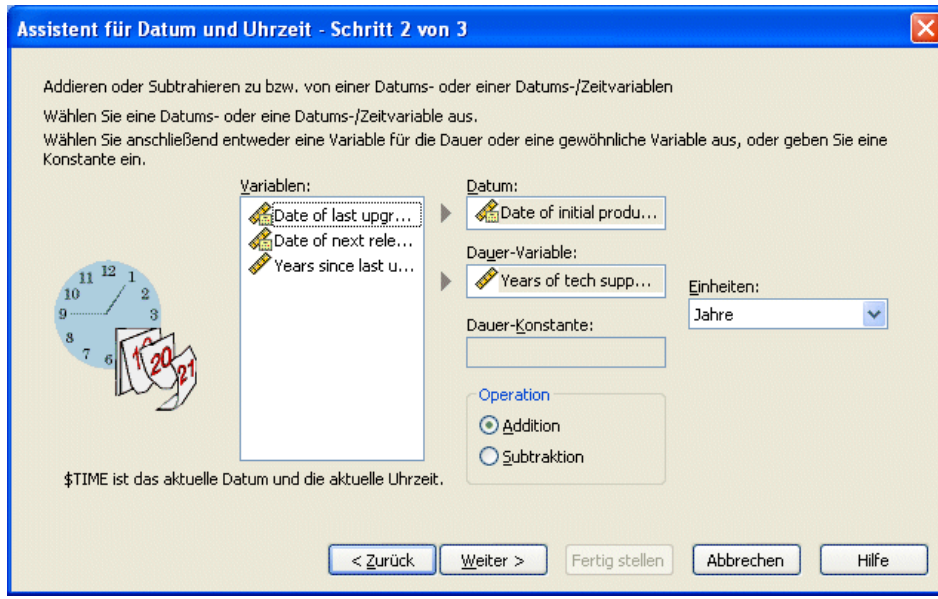
Addieren bzw. Subtrahieren einer Dauer zu bzw. von einem Datum

So addieren bzw. subtrahieren Sie eine Dauer zu bzw. von einer Variablen mit Datumsformat:

▶ Wählen Sie im Bildschirm "Durchführen von Berechnungen mit Datumswerten" des Assistenten für Datum und Uhrzeit die Option Addieren bzw. Subtrahieren einer Dauer zu bzw. von einem Datum aus.

Auswahl der Datums-/Zeitvariablen und der zu addierenden bzw. subtrahierenden Dauer

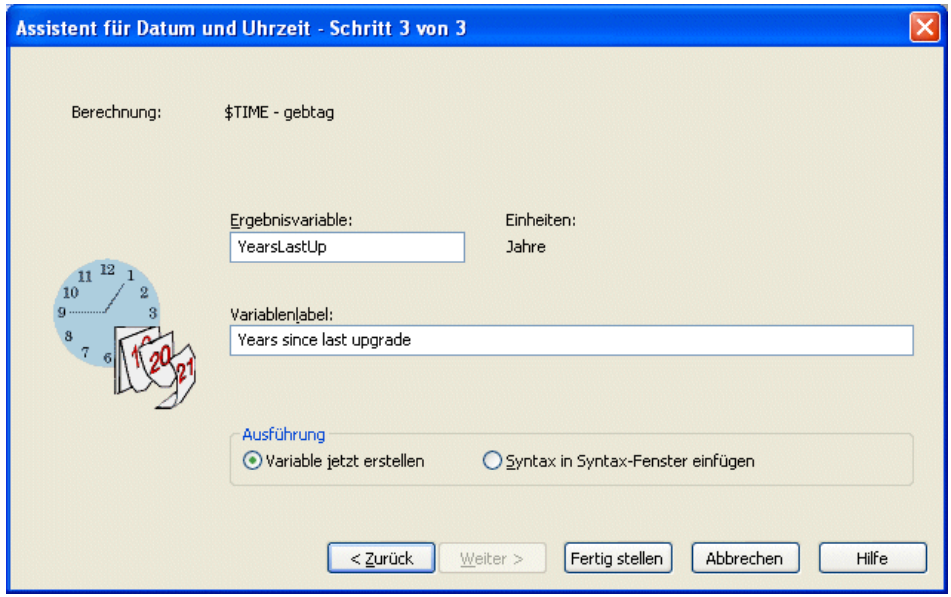
Abbildung 8-22
Addieren bzw. Subtrahieren einer Dauer, Schritt 2


Assistent für Datum und Uhrzeit - Schritt 2 von 3
Addieren oder Subtrahieren zu bzw. von einer Datums- oder einer Datums-/Zeitvariablen
Wählen Sie eine Datums- oder eine Datums-/Zeitvariable aus.
Wählen Sie anschließend entweder eine Variable für die Dauer oder eine gewöhnliche Variable aus, oder geben Sie eine Konstante ein.
Variablen:
Date of last upgr...
Date of next rele...
Years since last u...
Datum:
Date of initial produ...
Dauer-Variable:
Years of tech supp...
Einheiten:
Jahre
Dauer-Konstante:
Operation
Addition
Subtraktion
$TIME ist das aktuelle Datum und die aktuelle Uhrzeit.
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

- Wählen Sie eine Datums- oder Zeitvariable aus.
- Wählen Sie eine Dauer-Variable aus, oder geben Sie einen Wert für die Dauer-Konstante ein. Variablen, die für die Dauer verwendet werden, können keine Datums- bzw. Datums-/Zeitvariablen sein. Es kann sich bei ihnen um Dauer-Variablen oder einfache numerische Variablen handeln.
- Wählen Sie die Einheit für die Dauer aus der Dropdown-Liste aus. Wählen Sie Dauer aus, wenn Sie eine Variable verwenden und die Variable in einem Format für die Dauer verwenden, beispielsweise hh:mm oder hh:mm:ss.

Angeben des Ergebnisses der Addition bzw. Subtraktion einer Dauer zu bzw. von einer Datums-/Zeitvariablen

Abbildung 8-23
Addieren bzw. Subtrahieren einer Dauer, Schritt 3



- Geben Sie einen Namen für die Ergebnisvariable ein. Dieser darf nicht mit dem Namen einer bestehenden Variablen übereinstimmen.

  Die folgenden Optionen sind verfügbar:

  - Zuweisen eines beschreibenden Variablenlabels zur neuen Variablen.

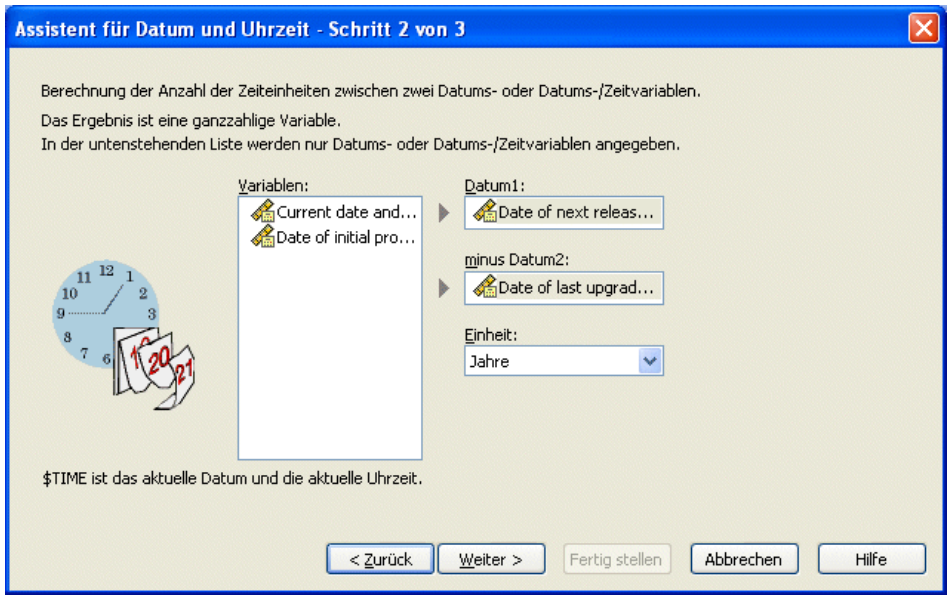
Subtrahieren von Variablen mit Datumsformat

So können Sie zwei Variablen mit Datumsformat subtrahieren:

- Wählen Sie im Bildschirm “Durchführen von Berechnungen mit Datumswerten” des Assistenten für Datum und Uhrzeit die Option Berechnen der Anzahl der Zeiteinheiten zwischen zwei Datumswerten aus.

Auswählen der zu subtrahierenden Variablen mit Datumsformat

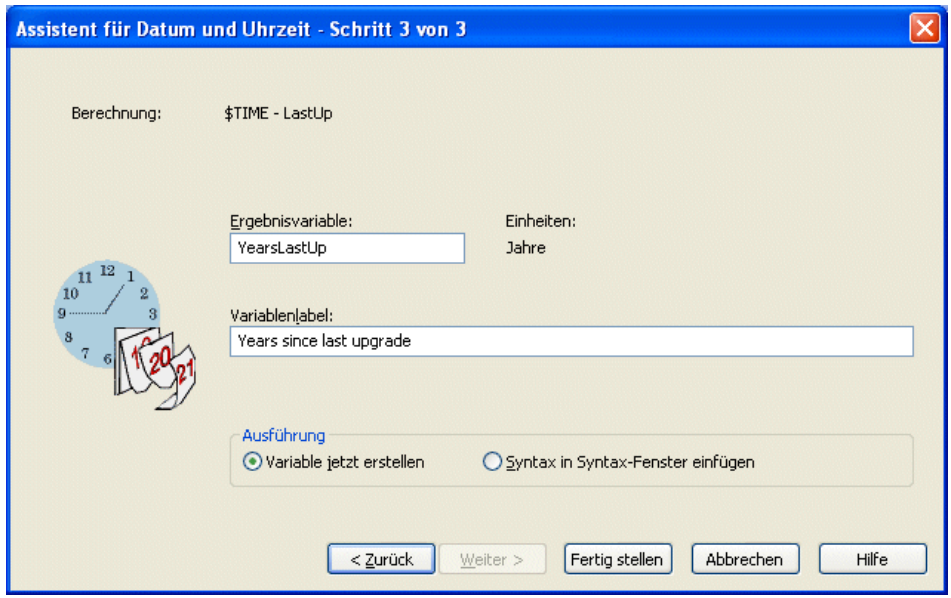
Abbildung 8-24
Subtrahieren von Datumswerten, Schritt 2


Assistent für Datum und Uhrzeit - Schritt 2 von 3
Berechnung der Anzahl der Zeiteinheiten zwischen zwei Datums- oder Datums-/Zeitvariablen.
Das Ergebnis ist eine ganzzahlige Variable.
In der untenstehenden Liste werden nur Datums- oder Datums-/Zeitvariablen angegeben.
Variablen:
Current date and...
Date of initial pro...
Datum1:
Date of next releas...
minus Datum2:
Date of last upgrad...
Einheit:
Jahre
$TIME ist das aktuelle Datum und die aktuelle Uhrzeit.
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

- Wählen Sie die zu subtrahierenden Variablen aus.
- Wählen Sie die Einheit für das Ergebnis aus der Dropdown-Liste aus.

Angeben des Ergebnisses der Subtraktion von zwei Variablen mit Datumsformat

Abbildung 8-25
Subtrahieren von Datumswerten, Schritt 3



- Geben Sie einen Namen für die Ergebnisvariable ein. Dieser darf nicht mit dem Namen einer bestehenden Variablen übereinstimmen.

  Die folgenden Optionen sind verfügbar:

  - Zuweisen eines beschreibenden Variablenlabels zur neuen Variablen.

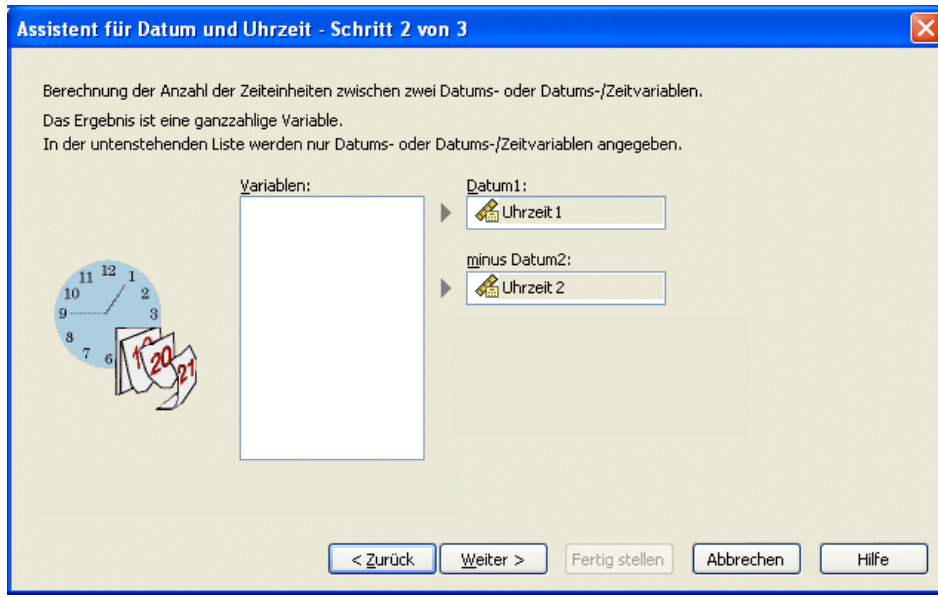
Subtrahieren von Dauer-Variablen

So können Sie zwei Dauer-Variablen subtrahieren:

- Wählen Sie im Bildschirm “Durchführen von Berechnungen mit Datumswerten” des Assistenten für Datum und Uhrzeit die Option Subtrahieren zweier Werte für Dauer aus.

Auswählen der zu subtrahierenden Dauer-Variablen

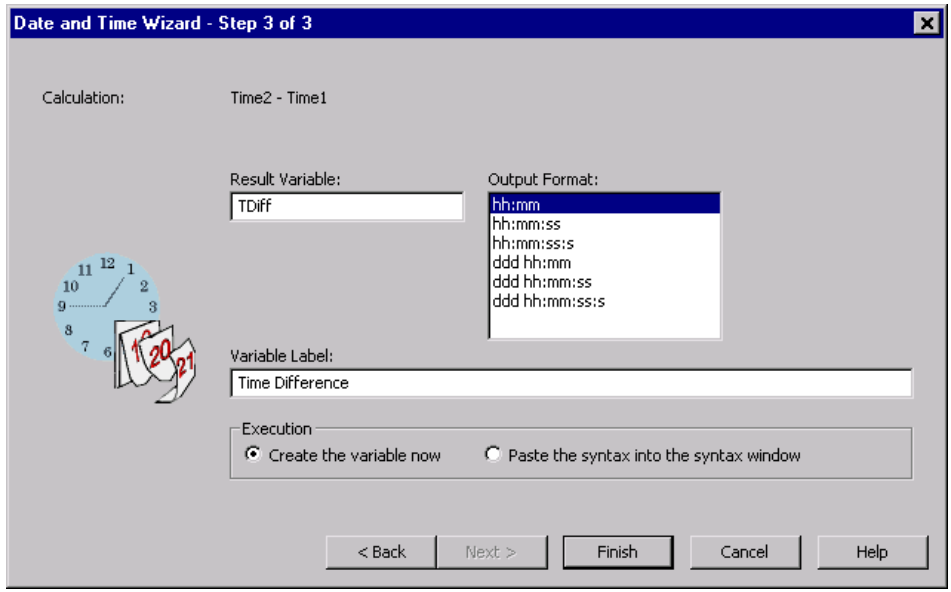
Abbildung 8-26
Subtrahieren von Werten für Dauer, Schritt 2


Assistent für Datum und Uhrzeit - Schritt 2 von 3
Berechnung der Anzahl der Zeiteinheiten zwischen zwei Datums- oder Datums-/Zeitvariablen.
Das Ergebnis ist eine ganzzahlige Variable.
In der untenstehenden Liste werden nur Datums- oder Datums-/Zeitvariablen angegeben.
Variablen:
Datum1:
Uhrzeit 1
minus Datum2:
Uhrzeit 2
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

▶ Wählen Sie die zu subtrahierenden Variablen aus.

Angeben des Ergebnisses der Subtraktion von zwei Dauer-Variablen

Abbildung 8-27
Subtrahieren von Werten für Dauer, Schritt 3



- Geben Sie einen Namen für die Ergebnisvariable ein. Dieser darf nicht mit dem Namen einer bestehenden Variablen übereinstimmen.
- Wählen Sie ein Format für die Dauer aus der Liste “Ausgabeformat” aus.

  Die folgenden Optionen sind verfügbar:

  - Zuweisen eines beschreibenden Variablenlabels zur neuen Variablen.

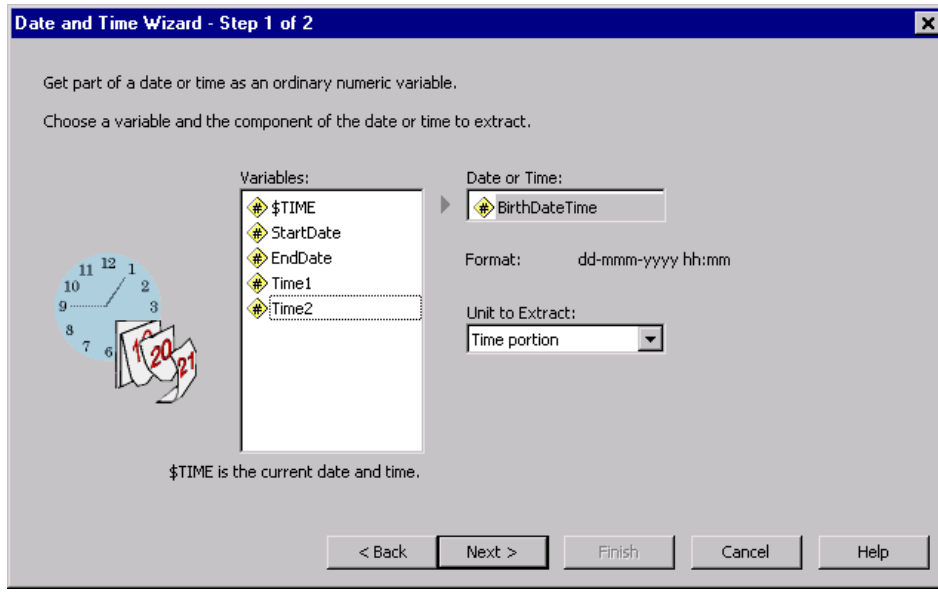
Extrahieren eines Teils einer Datums-/Zeitvariablen

So extrahieren Sie eine Komponente – beispielsweise das Jahr – aus einer Datums-/Zeitvariablen:

- Wählen Sie im Einführungsbildschirm des Assistenten für Datum und Uhrzeit die Option Einen Teil einer Datums- oder Zeitvariablen extrahieren aus.

Auswählen der aus der Datums-/Zeitvariablen zu extrahierenden Komponente

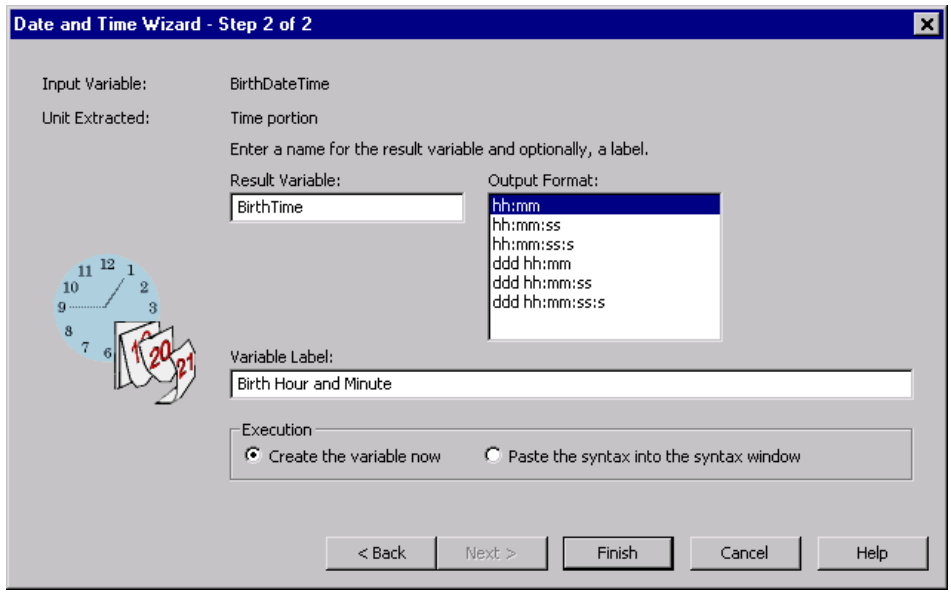
Abbildung 8-28
Abrufen eines Teils einer Datums-/Zeitvariablen, Schritt 1



- Wählen Sie die Variable aus, die den zu extrahierenden Datums- oder Zeitteil enthält.
- Wählen den zu extrahierenden Teil der Variablen aus der Dropdown-Liste aus. Sie können Informationen aus Datumsangaben extrahieren, die nicht explizit im angezeigten Datum enthalten sind, wie beispielsweise den Tag der Woche.

Angeben des Ergebnisses der Extraktion einer Komponente aus einer Datums-/Zeitvariablen

Abbildung 8-29
Abrufen eines Teils einer Datums-/Zeitvariablen, Schritt 2



- Geben Sie einen Namen für die Ergebnisvariable ein. Dieser darf nicht mit dem Namen einer bestehenden Variablen übereinstimmen.

- Beim Extrahieren des Datums- oder Zeitteils einer Datums-/Zeitvariablen müssen Sie in der Liste “Ausgabeformat” ein Format auswählen. Wenn kein Ausgabeformat erforderlich ist, ist die Liste “Ausgabeformat” deaktiviert.

  Die folgenden Optionen sind verfügbar:

  - Zuweisen eines beschreibenden Variablenlabels zur neuen Variablen.

Datentransformationen für Zeitreihen

In SPSS werden verschiedene Datentransformationen bereitgestellt, die bei Zeitreihenanalysen von Nutzen sind:

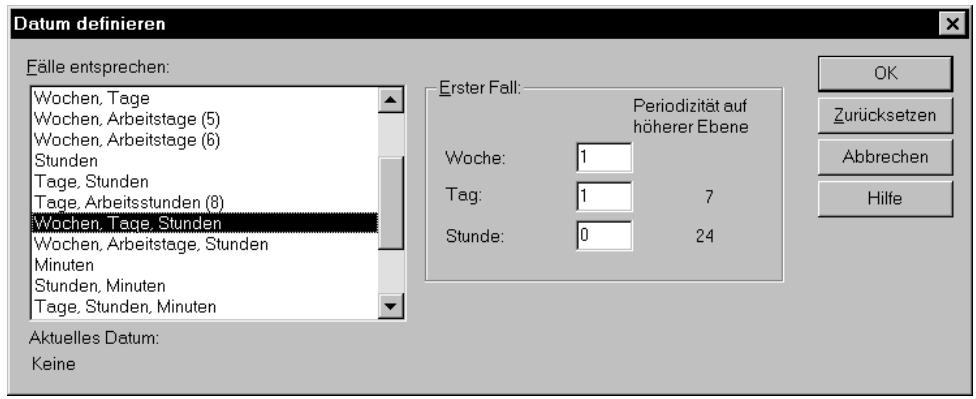
- Erzeugen von Datumsvariablen zum Erstellen von Periodizität und zum Unterscheiden zwischen historischen Perioden, Validierungsperioden und Vorhersageperioden,
- Erstellen von Zeitreihenvariablen als Funktionen vorhandener Zeitreihenvariablen,
- Ersetzen von system- und benutzerdefinierten fehlenden Werten durch Schätzwerte auf der Grundlage einer von mehreren möglichen Methoden.

Eine **Zeitreihe** wird erstellt, indem eine Variable (oder ein Variablen-Set) regelmäßig über einen Zeitraum beobachtet wird. Transformationen von Zeitreihendaten setzen eine Struktur in der Datendatei voraus, bei der jeder Fall (jede Zeile) eine Reihe von Beobachtungen zu einem unterschiedlichen Zeitpunkt darstellt und dabei die Zeitdauer zwischen den Fällen gleichförmig ist.

Datum definieren

Im Dialogfeld "Datum definieren" werden Datumsvariablen erstellt, die zum Herstellen der Periodizität einer **Zeitreihe** und zum Beschriften der Ausgabe aus Zeitreihenanalysen verwendet werden können.

Abbildung 8-30
Dialogfeld "Datum definieren"



Fälle entsprechen. Hiermit wird das zum Erstellen von Datumsangaben verwendete Zeitintervall definiert.

- Mit Kein Datum werden alle bisher definierten Datumsvariablen entfernt. Dabei werden Variablen mit den folgenden Namen gelöscht: *Jahr_*, *Quartal_*, *Monat_*, *Woche_*, *Tag_*, *Stunde_*, *Minute_*, *Sekunde_* und *Datum_*.
- Mit Benutzerdefiniert wird das Vorhandensein benutzerdefinierter Datumsvariablen angezeigt, die mit der Befehlssyntax erstellt wurden, zum Beispiel eine viertägige Arbeitswoche. Dieser Eintrag spiegelt nur den aktuellen Stand der Arbeitsdatei wider. Die Auswahl dieses Eintrags in der Liste hat keine Auswirkung.

Erster Fall. Hiermit wird der Wert für das Startdatum definiert, das dem ersten Fall zugeordnet ist. Nachfolgenden Fällen werden auf dem Zeitintervall basierende sequentielle Werte zugeordnet.

Periodizität auf höherer Ebene. Hier wird die wiederholte zyklische Schwankung angezeigt, wie zum Beispiel die Anzahl der Monate in einem Jahr oder die Anzahl der Tage in einer Woche. Der angezeigte Wert ist der höchste Wert, den Sie eingeben können.

Für jede zum Definieren des Datums verwendete Komponente wird eine neue numerische Variable erzeugt. Die neuen Namen der Variablen enden mit einem Unterstrich. Es wird außerdem die beschreibende String-Variable *Datum_* aus den Komponenten erzeugt. Wenn Sie zum Beispiel Wochen, Tage, Stunden ausgewählt haben, werden vier neue Variablen erstellt: *Woche_*, *Tag_*, *Stunde_* und *Datum_*.

Wenn bereits Datumsvariablen definiert wurden, werden diese beim Definieren neuer Datumsvariablen mit gleichem Namen ersetzt.

So definieren Sie Datumsangaben für Zeitreihendaten:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Daten
  Datum definieren...

▶ Wählen Sie ein Zeitintervall aus der Liste “Fälle entsprechen:” aus.

▶ Geben Sie die Werte ein, die das Startdatum für “Erster Fall” definieren. Hiermit wird das dem ersten Fall zugewiesene Datum bestimmt.

Vergleich von Datumsvariablen und Variablen im Datumsformat

Mit “Datum definieren” erzeugte Datumsvariablen dürfen nicht mit Variablen im Datumsformat verwechselt werden, die in der Variablenansicht des Daten-Editors definiert werden. Datumsvariablen werden verwendet, um Periodizität für Zeitreihendaten zu erstellen. Variablen im Datumsformat stellen in verschiedenen Datums- und Uhrzeitformaten angezeigte Datums- und Uhrzeitangaben dar. Datumsvariablen sind einfache ganze Zahlen, welche die Anzahl von Tagen, Wochen, Stunden usw. ab einem benutzerdefinierten Ausgangspunkt angeben. Intern werden die meisten Variablen im Datumsformat als die Anzahl der seit dem 14. Oktober 1582 vergangenen Sekunden gespeichert.

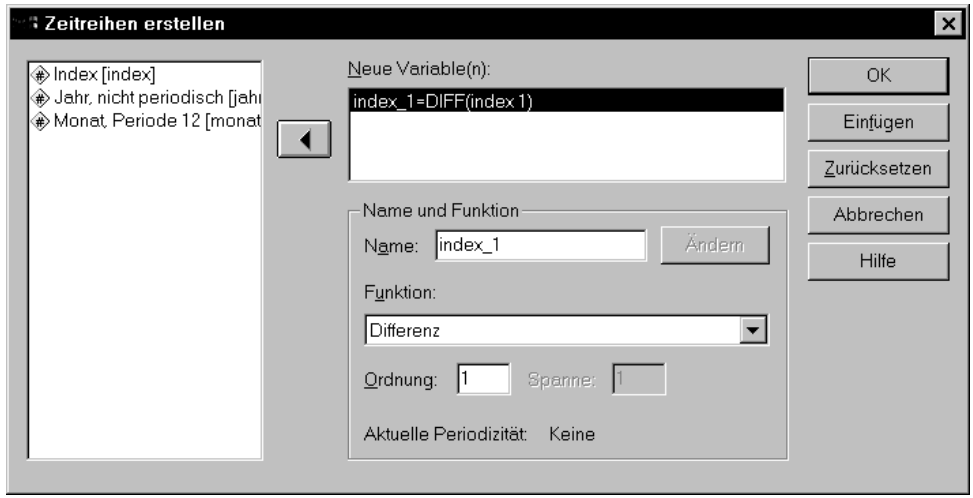
Zeitreihen erstellen

Im Dialogfeld “Zeitreihen erstellen” werden neue Variablen auf der Grundlage der Funktionen von vorhandenen numerischen **Zeitreihenvariablen** erstellt. Diese transformierten Werte werden in vielen Prozeduren zur Zeitreihenanalyse benutzt.

Neue Variablennamen bestehen in der Standardeinstellung aus den ersten sechs Zeichen der vorhandenen Variablen, aus denen sie erstellt wurden, einem Unterstrich und einer laufenden Nummer. Der neue Variablenname für die Variable *Preis* ist z. B. *Preis_1*. Den neuen Variablen werden alle definierten Wertelabels der ursprünglichen Variablen zugewiesen.

Zu den verfügbaren Funktionen zum Erzeugen von Zeitreihenvariablen gehören Differenzen, gleitende Durchschnitte, gleitende Mediane, Intervall- und Vorlauffunktionen.

Abbildung 8-31
Dialogfeld "Zeitreihen erstellen"



So erstellen Sie eine neue Zeitreihenvariable:

- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

  Transformieren
  Zeitreihen erstellen...

- ▶ Wählen Sie die Zeitreihenfunktion, die Sie zum Transformieren der ursprünglichen Variablen verwenden möchten.

- ▶ Wählen Sie die Variablen, aus denen Sie neue Zeitreihenvariablen erstellen möchten. Es können nur numerische Variablen verwendet werden.

  Die folgenden Optionen sind verfügbar:

  - Durch Eingabe neuer Variablennamen können die vorgegebenen Variablennamen überschrieben werden.
  - Ändern der Funktion für eine ausgewählte Variable.

Funktionen zur Transformation von Zeitreihen

Differenz. Die nichtsaisonale Differenz zwischen aufeinanderfolgenden Werten in den Datenreihen. Die Ordnung ist die Anzahl der zum Berechnen der Differenz verwendeten zurückliegenden Werte. Da für jede Ordnung der Differenzen eine Beobachtung fehlt, sind systemdefinierte fehlende Werte am Anfang der Datenreihe vorhanden. Wenn die Ordnung der Differenzen zum Beispiel 2 beträgt, besitzen die ersten zwei Fälle den systemdefinierten fehlenden Wert in der neuen Variablen.

Saisonale Differenz. Differenz zwischen Reihenwerten, die eine konstante Spanne auseinanderliegen. Die Spanne basiert auf der aktuell definierten Periodizität. Zum Berechnen saisonaler Differenzen müssen Datumsvariablen (Menü "Daten", Befehl "Datum definieren") mit einer periodischen Komponente (wie den Monaten eines Jahres) definiert sein. Die Ordnung ist die Anzahl der zum Berechnen der Differenz verwendeten saisonalen Perioden. Die Anzahl der Fälle mit dem systemdefinierten fehlenden Wert am Anfang der Datenreihen ist gleich dem Produkt aus der Periodizität und der Ordnung. Wenn zum Beispiel die aktuelle Periodizität 12 und die Reihenfolge 2 beträgt, besitzen die ersten 24 Fälle den systemdefinierten fehlenden Wert als neue Variable.

Zentrierter gleitender Durchschnitt. Durchschnitt einer Spanne von Datenreihenwerten, die den aktuellen Wert umgeben und einschließen. Die Spanne ist die Anzahl der zum Berechnen des Durchschnitts verwendeten Datenreihenwerte. Wenn der Wert der Spanne gerade ist, wird der gleitende Durchschnitt so berechnet, daß für jedes Paar nichtzentrierter Mittelwerte der Durchschnitt gebildet wird. Die Anzahl der Fälle mit dem systemdefinierten fehlenden Wert am Anfang und Ende der Datenreihe für eine Spanne von n ist gleich $n/2$ bei geraden und ungeraden Werten für die Spanne. Wenn die Spanne zum Beispiel 5 beträgt, gibt es 2 Fälle mit dem systemdefinierten fehlenden Wert am Anfang und am Ende der Datenreihe.

Zurückgreifender gleitender Durchschnitt. Durchschnitt der Spanne von Datenreihenwerten vor dem aktuellen Wert. Die Spanne ist die Anzahl der zum Berechnen des Durchschnitts verwendeten vorangehenden Datenreihenwerte. Die Anzahl der Fälle mit dem systemdefinierten fehlenden Wert am Anfang der Datenreihe ist gleich dem Wert der Spanne.

Gleitende Mediane. Median einer Spanne von Datenreihenwerten, die den aktuellen Wert umgeben und einschließen. Die Spanne ist die Anzahl der zum Berechnen des Medians verwendeten Datenreihenwerte. Wenn die Spanne geradzahlig ist, wird der Median durch Berechnen des Durchschnitts jedes Paars unzentrierter Mediane

ermittelt. Die Anzahl der Fälle mit dem systemdefinierten fehlenden Wert am Anfang und Ende der Datenreihe für eine Spanne von *n* ist gleich *n*/2 bei geraden und ungeraden Werten für die Spanne. Wenn die Spanne zum Beispiel 5 beträgt, gibt es 2 Fälle mit dem systemdefinierten fehlenden Wert am Anfang und am Ende der Datenreihe.

Kumulierte Summe. Kumulierte Summe der Datenreihenwerte bis zum und einschließlich des aktuellen Werts.

Lag. Wert eines zurückliegenden Falls auf der Grundlage der angegebenen Ordnung der Intervalle. Die Ordnung ist die Anzahl der Fälle vor dem gegenwärtigen Fall, aus dem der Wert ermittelt wird. Die Anzahl der Fälle mit dem systemdefinierten fehlenden Wert am Anfang der Datenreihe ist gleich der Ordnung.

Lead. Wert eines nachfolgenden Falls auf der Grundlage der angegebenen Ordnung der Intervalle. Die Ordnung ist die Anzahl der Fälle nach dem aktuellen Fall, aus dem der Wert ermittelt wird. Die Anzahl der Fälle mit dem systemdefinierten fehlenden Wert am Ende der Datenreihe ist gleich der Ordnung.

Glätten. Neue Datenreihenwerte auf der Grundlage einer Glättung von zusammengesetzten Daten. Die Glättung beginnt mit einem gleitenden Median von 4, der von einem gleitenden Median von 2 zentriert wird. Dann werden die Werte erneut durch Anwendung eines gleitenden Medians von 5, eines gleitenden Medians von 3 und Hanning (gleitende gewichtete Mittelwerte) geglättet. Residuen werden durch Subtrahieren der geglätteten Datenreihen von den ursprünglichen Datenreihen berechnet. Dieser vollständige Prozeß wird dann erneut auf die errechneten Residuen angewendet. Zuletzt werden durch Subtrahieren der beim ersten Durchlauf dieses Prozesses errechneten Werte die geglätteten Residuen ermittelt. Dieses Verfahren wird auch als T4253H-Glätten bezeichnet.

Fehlende Werte ersetzen

Fehlende Beobachtungen können Probleme in der Analyse aufwerfen, und einige Maße für Zeitreihen können bei fehlenden Werten in den Datenreihen nicht berechnet werden. In einigen Fällen ist der Wert für eine bestimmte Beobachtung einfach nicht bekannt. Fehlende Daten können auch aus den folgenden Ursachen entstehen:

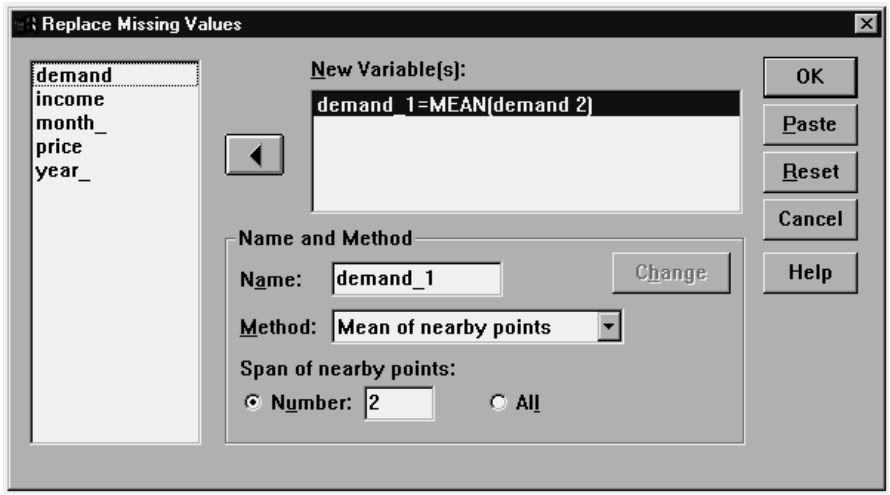
- Jeder Grad der Differenzierung verkürzt eine Reihe um ein Element.
- Jeder Grad der saisonalen Differenzierung verkürzt eine Reihe um eine Saison.

- Wenn Sie eine neue Reihe mit Prognosen erstellen, die über das Ende der vorhandenen Reihe hinausreichen (indem Sie auf die Schaltfläche Speichern klicken und eine geeignete Auswahl treffen), weisen die ursprüngliche Reihe und die erzeugte Residuenreihe fehlende Werte für die neuen Beobachtungen auf.
- Einige Transformationen (z. B. die Log-Transformation) erzeugen fehlende Daten für bestimmte Werte in der ursprünglichen Reihe.

Fehlende Daten am Anfang oder Ende einer Zeitreihe stellen kein größeres Problem dar. Sie verkürzen nur die brauchbare Länge der Zeitreihe. Lücken im Inneren einer Zeitreihe (*eingebettete* fehlende Daten) können ein viel schwerwiegenderes Problem darstellen. Das Ausmaß des Problems ist abhängig vom eingesetzten Analyseverfahren.

Im Dialogfeld "Fehlende Werte ersetzen" werden neue **Zeitreihenvariablen** aus bereits vorhandenen erstellt. Dabei werden fehlende Werte durch Schätzwerte ersetzt, die mit einer von mehreren möglichen Methoden errechnet werden. Neue Variablennamen bestehen in der Standardeinstellung aus den ersten sechs Zeichen der vorhandenen Variablen, aus denen sie erstellt wurden, einem Unterstrich und einer laufenden Nummer. Der neue Variablenname für die Variable *Preis* ist z. B. *Preis_1*. Den neuen Variablen werden alle definierten Wertelabels der ursprünglichen Variablen zugewiesen.

Abbildung 8-32
Dialogfeld "Fehlende Werte ersetzen"



So ersetzen Sie fehlende Werte für Zeitreihenvariablen:

- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Transformieren
  Fehlende Werte ersetzen...
- Wählen Sie die zum Ersetzen fehlender Werte zu verwendende Schätzmethode aus.
- Wählen Sie die Variablen, für die Sie fehlende Werte ersetzen möchten.

  Die folgenden Optionen sind verfügbar:

  - Durch Eingabe neuer Variablennamen können die vorgegebenen Variablennamen überschrieben werden.
  - Die Schätzmethode für eine ausgewählte Variable kann geändert werden.

Schätzmethoden zum Ersetzen fehlender Werte

Zeitreihen-Mittelwert. Fehlende Werte werden durch den Mittelwert der gesamten Datenreihe ersetzt.

Mittel der Nachbarpunkte. Fehlende Werte werden durch den Mittelwert gültiger Umgebungswerte ersetzt. Die Anzahl der Nachbarpunkte ist die Anzahl gültiger Werte über und unter dem zum Berechnen des Mittelwerts verwendeten fehlenden Wert.

Median der Nachbarpunkte. Fehlende Werte werden mit dem Median gültiger Umgebungswerte ersetzt. Die Anzahl der Nachbarpunkte ist die Anzahl gültiger Werte über und unter dem zum Berechnen des Medians verwendeten fehlenden Wert.

Lineare Interpolation. Fehlende Werte werden unter Anwendung einer linearen Interpolation ersetzt. Für die Interpolation werden der letzte gültige Wert vor dem fehlenden Wert und der erste gültige Wert nach dem fehlenden Wert verwendet. Wenn der erste oder letzte Fall in der Datenreihe einen fehlenden Wert hat, wird der fehlende Wert nicht ersetzt.

Linearer Trend am Punkt. Fehlende Werte werden mit dem linearen Trend für den Punkt ersetzt. Für die vorhandene Datenreihe wird eine Regression auf eine von 1 bis *n* skalierte Indexvariable ausgeführt. Fehlende Werte werden mit ihren vorhergesagten Werten ersetzt.

Bewerten von Daten mit Vorhersagemodellen

Die Anwendung eines Vorhersagemodells auf eine Datenmenge wird als **Bewertung** (Scoring) der Daten bezeichnet. SPSS, Clementine und AnswerTree bieten Verfahren für den Aufbau von Vorhersagemodellen, wie Regression, Clustern, sowie Baummodelle und Modelle für neurale Netzwerke. Sobald ein Modell erstellt wurde, können die Modellspezifikationen als XML-Datei mit allen für die Rekonstruktion des Modells erforderlichen Informationen gespeichert werden. Das SPSS Server-Produkt bietet anschließend die Möglichkeit, eine XML-Modelldatei zu lesen und das Modell auf ein Daten-Set anzuwenden.

Beispiel. Eine Kreditanwendung wird auf der Grundlage verschiedener Aspekte des Bewerbers und des betreffenden Kredits nach Risiko bewertet. Anhand des aus dem Risikomodell gewonnenen Score-Wertes für den Kredit wird die Kreditanwendung angenommen oder zurückgewiesen.

Die Bewertung (Scoring) wird als Datentransformation behandelt. Das Modell wird intern als Menge numerischer Transformationen ausgedrückt, die auf ein vorgegebenes Variablen-Set (die im Modell festgelegten Einflussvariablen) angewendet werden, um ein vorhergesagtes Ergebnis zu erzielen. In diesem Sinne ist der Vorgang der Bewertung von Daten mit Hilfe eines bestimmten Modells im Grunde dasselbe wie die Anwendung jeder anderen Funktion, wie beispielsweise der Funktion “Quadratwurzel”, auf eine Datenmenge.

Die Bewertung ist nur mit SPSS Server verfügbar und kann von Benutzern interaktiv ausgeführt werden, die im Modus für verteilte Analysen arbeiten. Bei der Bewertung von großen Datendateien ist die Verwendung von SPSS Batch Facility zu empfehlen. Dies ist eine gesonderte, ausführbaren Version von SPSS, die im Lieferumfang von SPSS Server enthalten ist. Informationen zur Verwendung von SPSS Batch Facility finden Sie imBenutzerhandbuch zu SPSS Batch Facility, das Sie als PDF-Datei auf der Produkt-CD von SPSS Server finden.

Der Bewertungsprozess besteht aus den folgenden Schritten:

- Laden eines Modells aus einer Datei im XML- (PMML-) Format.
- Berechnen der Werte als neue Variable unter Verwendung von ApplyModel oder StrApplyModel im Dialogfeld Variable berechnen .

Weitere Informationen über die Funktionen `ApplyModel` und `StrApplyModel` finden Sie im Handbuch *SPSS Command Syntax Reference* im Abschnitt "Transformation Expressions" unter "Scoring Expressions".

Laden eines gespeicherten Modells

Mithilfe des Dialogfelds "Modell laden" können Sie Vorhersagemodelle laden, die im XML- (PMML-) Format gespeichert wurden. Das Dialogfeld ist nur verfügbar, wenn Sie im Modus für verteilte Analysen arbeiten. Sie müssen ein Modell laden, ehe Sie damit Daten bewerten können.

Abbildung 8-33
Ausgabe beim Laden eines Modells

mregression

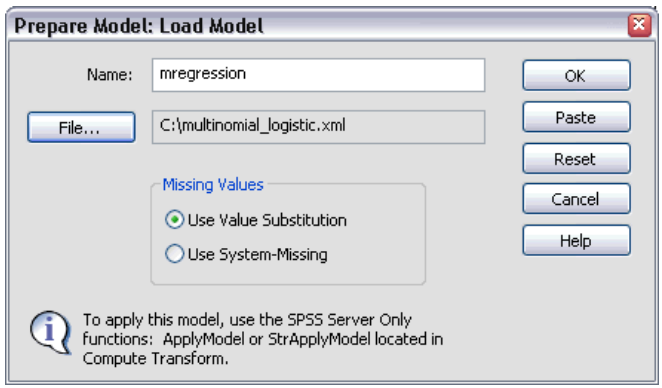
	Model Variables					
SPSS Variables	Name	Label	Type	Width	Role	Measurement
Age_Group	Age_ Group	Age_ Group	Numeric	8	Factor	Nominal
Frequency	Frequency	Frequency	Numeric	7	Covariate	Continuous
Gender	Gender	Gender	String	6	Factor	Nominal
Has_Broadband	Has_ Broadband	Has_ Broadband	Numeric	10	Factor	Nominal
Has_Child	Has_Child	Has_Child	Numeric	10	Factor	Nominal
Income_Group	Income_ Group	Income_ Group	Numeric	10	Factor	Nominal
Log_Amount	Log_ Amount	Log_ Amount	Numeric	8	Covariate	Continuous
Recency	Recency	Recency	Numeric	10	Covariate	Continuous
Response	Response	Response	Numeric	10	Target	Nominal

So laden Sie ein Modell

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Transformieren
  Modell vorbereiten
    Modell laden...

Abbildung 8-34
Dialogfeld "Modell vorbereiten: Modell laden"



- Geben Sie einen Namen ein, der diesem Modell zugeordnet werden soll. Jedes geladenen Modell muss einen eindeutigen Namen aufweisen.

- Klicken Sie auf Datei, und wählen Sie eine Modelldatei aus. Im dadurch geöffneten Dialogfeld "Datei öffnen" werden die Dateien angezeigt, die im Modus für verteilte Analysen verfügbar sind. Hierbei handelt es sich um die Dateien auf dem Computer, auf dem der SPSS Server installiert wurde, und um Dateien in freigegebenen Ordnen oder Laufwerken auf Ihrem lokalen Computer.

  Anmerkung: Wenn Sie Daten bewerten, wird das Modell auf Variablen in der Arbeitsdatei mit denselben Namen wie die Variablen aus der Modelldatei angewendet. Sie können Variablen aus dem ursprünglichen Modell anderen Variablen in der Arbeitsdatei zuordnen, indem Sie die Befehlssyntax verwenden (Informationen hierzu finden unter dem Befehl `MODEL HANDLE`).

  Name. Ein Name zum Identifizieren dieses Modells. Es gelten dieselben Regeln für gültige Modellnamen wie für SPSS-Variablennamen (siehe "Variablennamen" in Kapitel 5 auf S. 99), wobei zusätzlich das Zeichen $ als erstes Zeichen zulässig ist. Sie verwenden diesen Namen, um das Modell anzugeben, wenn Sie die Daten mit den Funktionen ApplyModel oder StrApplyModel bewerten.

  Datei. Die XML- (PMML-) Datei, die die Modellspezifikation enthält.

Fehlende Werte

In diesem Gruppenfeld legen Sie die Behandlung von fehlenden Werten für die im Modell definierten Einflussvariablen fest, die während des Bewertungsprozesses auftreten. Bei der Bewertung liegt in den folgenden Fällen ein fehlender Wert vor:

- Eine Einflussvariable enthält keinen Wert. Bei numerischen Variablen entspricht dies dem systemdefinierten fehlenden Wert. Bei String-Variablen entspricht dies einem leeren String.
- Der Wert für die vorliegende Einflussvariable wurde im Modell als benutzerdefiniert fehlend definiert. Werte, die in der Arbeitsdatei, aber nicht im Modell, als benutzerdefiniert fehlend definiert sind, werden im Bewertungsprozess nicht als fehlende Werte behandelt.
- Die Einflussvariable ist kategorial und der Wert entspricht keiner der im Modell definierten Kategorien.

Werte ersetzen. Hierbei wird versucht, fehlende Werte beim Bewerten von Fällen zu ersetzen. Die Methode für das Bestimmen eines Werts, der einen fehlenden Wert ersetzen soll, hängt von der Art des Vorhersagemodells ab.

- **SPSS-Modelle.** Wenn beim Erstellen und Speichern eines linearen Regressionsmodells oder eines Diskriminanzmodells für die unabhängigen Variablen festgelegt wurde, dass fehlende Werte durch den Mittelwert ersetzt werden sollen, wird bei der Berechnung der Bewertung dieser Mittelwert anstelle des fehlenden Werts verwendet. Wenn der Mittelwert nicht verfügbar ist, wird der systemdefinierte fehlende Wert zurückgegeben.
- **AnswerTree-Modelle und Modelle des SPSS-Befehls TREE.** Bei CHAID- und Exhaustive CHAID-Modellen wird für eine fehlende Teilungsvariable der größte untergeordnete Knoten ausgewählt. Der größte untergeordnete Knoten ist der Knoten mit der größten Grundgesamtheit unter den untergeordneten Knoten auf der Grundlage einer Stichprobe von Lernfällen. Für C&RT- und QUEST-Modelle werden (wenn überhaupt) zuerst Ersatzteilungsvariablen verwendet. (Ersatzteilungsvariablen sind Teilungsvariablen, mit denen versucht wird, anhand von anderen Einflussvariablen eine möglichst starke Übereinstimmung mit der ursprünglichen Teilung zu erzielen.) Wenn keine Ersatzteilungsvariablen angegeben werden oder alle Ersatzteilungsvariablen fehlen, wird der größte untergeordnete Knoten verwendet.

- **Clementine-Modelle.** Lineare Regressionsmodelle werden behandelt wie im Abschnitt "SPSS-Modelle" erläutert. Logistische Regressionsmodelle werden behandelt, wie unter "Logistische Regressionsmodelle" beschrieben. C&RT Tree-Modelle werden behandelt, wie im Abschnitt zu C&RT-Modellen unter "AnswerTree-Modelle" beschrieben.
- **Logistische Regressionsmodelle.** Wenn bei Kovariaten in logistischen Regressionsmodellen ein Mittelwert der Einflussvariablen als Teil des gespeicherten Modells aufgenommen wurde, wird bei der Berechnung der Bewertung dieser Mittelwert anstelle des fehlenden Werts verwendet. Wenn die Einflussvariable kategorial ist (z. B. ein Faktor in einem logistischen Regressionsmodell), oder wenn der Mittelwert nicht verfügbar ist, wird der systemdefinierte fehlende Wert zurückgegeben.

Systemdefinierte fehlende Werte verwenden. Beim Bewerten eines Falls mit einem fehlenden Wert wird der systemdefinierte fehlende Wert zurückgegeben.

Anzeigen einer Liste der geladenen Modelle

Sie können eine Liste der aktuell geladenen Modelle erhalten. Wählen Sie die folgenden Befehle aus den Menüs aus (nur im Modus für verteilte Analysen verfügbar):

Transformieren
  Modell vorbereiten
    Modell(e) auflisten

Hiermit wird eine Tabelle mit den Modell-Bezeichnungen erstellt. Die Tabelle enthält eine Liste aller aktuell geladenen Modelle sowie für jedes Modell den zugeordneten Namen (den so genannten Modell-Bezeichnungen), den Modelltyp, den Pfad für die Modelldatei und die Methode zur Behandlung fehlender Werte.

Abbildung 8-35
Liste der geladenen Modelle

Model Handles

Name	Type	Input File Name	Validation handling	Missing handling
mregression	NOMREG	C:\models\mregression.xml	ON	SUBSTITUTE
tree	TREE	C:\models\tree.xml	ON	SUBSTITUTE

Zusätzliche Funktionen bei der Befehlssyntax

Sie können die ausgewählten Optionen im Dialogfeld “Modell laden” in ein Syntaxfenster einfügen und die entsprechende Befehlssyntax von `MODEL HANDLE` bearbeiten. Hierbei haben Sie folgende Möglichkeiten:

- Sie können Variablen aus dem ursprünglichen Modell anderen Variablen in der Arbeitsdatei zuordnen (mit dem Unterbefehl `MAP`). In der Standardeinstellung wird das Modell auf Variablen in der Arbeitsdatei mit denselben Namen wie die Variablen aus der Modelldatei angewendet.

Vollständige Informationen zur Syntax finden Sie im *SPSS Syntax Reference Guide*.

Kapitel
9

Umgang mit Dateien und Dateitransformationen

Datendateien liegen nicht immer genau in der Form vor, die Sie gerade benötigen. Vielleicht möchten Sie Datendateien zusammenfügen, die Daten in einer anderen Reihenfolge sortieren lassen, eine Teilmenge von Fällen auswählen oder die Einheit für die Analyse durch Gruppieren von Fällen ändern. In SPSS stehen Ihnen mehrere Möglichkeiten zum Transponieren von Dateien zur Verfügung, darunter folgende Funktionen:

Sortieren von Daten. Sie können Fälle nach dem Wert einer oder mehrerer Variablen sortieren lassen.

Transponieren von Fällen und Variablen. Das Format für SPSS-Datendateien ist so definiert, daß Zeilen als Fälle und Spalten als Variablen eingelesen werden. Bei Datendateien, in denen diese Reihenfolge umgekehrt ist, können Sie die Zeilen und Spalten vertauschen und so die Daten im richtigen Format einlesen.

Zusammenfügen von Dateien. Sie können zwei oder mehr Datendateien zusammenfügen. Es können Dateien zusammengefügt werden, welche dieselben Variablen, aber verschiedene Fälle enthalten, oder Dateien mit denselben Fällen und unterschiedlichen Variablen.

Auswählen von Teilmengen von Fällen. Sie können die Analyse auf eine Teilmenge von Fällen beschränken oder Analysen für verschiedene Teilmengen gleichzeitig vornehmen.

Aggregieren von Daten. Sie können die Einheit für die Analyse ändern, indem Sie Fälle nach Wert(en) einer oder mehrerer Gruppenvariablen aggregieren.

Gewichten von Daten. Sie können Fälle für die Analyse nach dem Wert einer Gewichtungsvariablen gewichten.

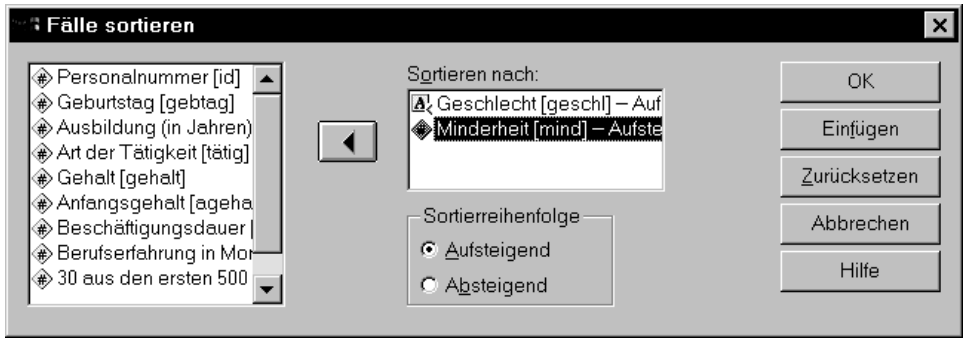
Umstrukturieren von Daten. Sie können Daten umstrukturieren und somit einen Fall (Datensatz) aus mehreren Fällen bzw. mehrere Fälle aus einem Fall erstellen.

Fälle sortieren

Mit diesem Dialogfeld können Sie Fälle (also Zeilen) der Datendatei nach einer oder mehreren Sortiervariablen sortieren. Sie können Fälle in aufsteigender oder absteigender Folge sortieren.

- Wenn Sie mehrere Sortiervariablen auswählen, werden die Fälle nach Variablen in den Kategorien der vorhergehenden Variablen aus der Liste "Sortieren nach" sortiert. Wenn Sie zum Beispiel *Geschl* als die erste Sortiervariable und *mind* als die zweite auswählen, wird innerhalb der Geschlechtskategorien nach Minderheit sortiert.
- Bei String-Variablen werden Großbuchstaben vor Kleinbuchstaben eingeordnet. So kommt zum Beispiel der String-Wert "Ja" in der Sortierfolge vor "ja".

Abbildung 9-1
Dialogfeld "Fälle sortieren"



So sortieren Sie Fälle:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Daten
  Fälle sortieren

▶ Wählen Sie mindestens eine Sortiervariable aus.

Transponieren

Mit “Transponieren” können Sie eine neue Datendatei anlegen, in welcher die Zeilen und Spalten aus der ursprünglichen Datendatei transponiert wurden, so daß die Fälle (also die Zeilen) zu Variablen und die Variablen (also die Spalten) zu Fällen werden. SPSS erzeugt automatisch neue Variablennamen und zeigt eine Liste der neuen Variablennamen an.

- SPSS erzeugt automatisch eine neue String-Variable, *case_lbl*, welche den ursprünglichen Variablennamen enthält.
- Falls die Arbeitsdatei eine ID- oder Namensvariable mit eindeutigen Werten enthält, können Sie diese im Feld “Namensvariable” eintragen. Die Werte dieser Variablen werden dann als Variablennamen in der transponierten Datendatei verwendet. Falls es sich um eine numerische Variable handelt, beginnen die Variablennamen mit dem Buchstaben *V*, gefolgt von einem numerischen Wert.
- Benutzerdefinierte fehlende Werte werden in der transponierten Datendatei in systemdefinierte fehlende Werte umgewandelt. Um diese Werte beizubehalten, müssen Sie die Definition fehlender Werte in der Variablenansicht des Daten-Editors ändern.

So transponieren Sie Variablen und Fälle:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Daten
  Transponieren...

▶ Wählen Sie eine oder mehrere Variablen aus, die in Fälle umgewandelt werden sollen.

Zusammenfügen von Datendateien

Mit SPSS können Sie Daten aus zwei Dateien auf zwei verschiedene Arten zusammenfügen. Sie verfügen über folgende Möglichkeiten:

- Zusammenfügen der Arbeitsdatei mit einem weiteren geöffneten Daten-Set oder einer Datendatei im SPSS-Format, das bzw. die dieselben Variablen aber unterschiedliche Fälle enthält.
- Zusammenfügen der Arbeitsdatei mit einem weiteren geöffneten Daten-Set oder einer Datendatei im SPSS-Format, das bzw. die dieselben Fälle aber unterschiedliche Variablen enthält.

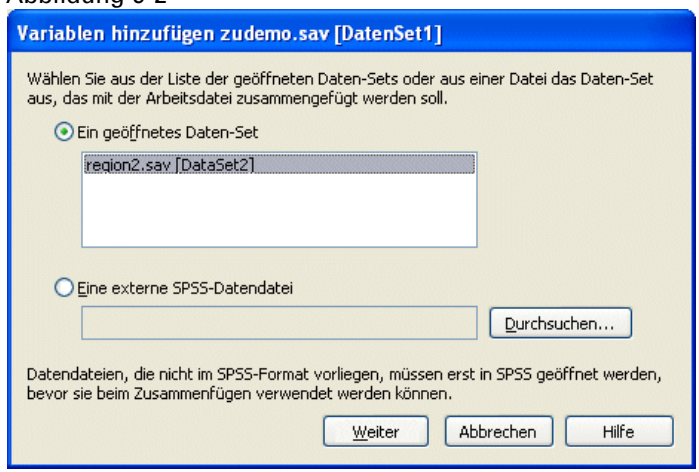
So fügen Sie Dateien zusammen

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Daten
  Dateien zusammenfügen

- Wählen Sie Fälle hinzufügen oder Variablen hinzufügen aus.

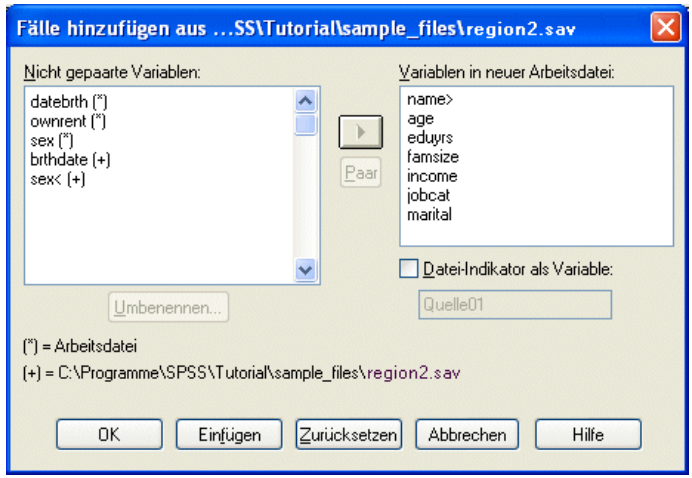
Abbildung 9-2


Variablen hinzufügen zudemo.sav [DatenSet1]
Wählen Sie aus der Liste der geöffneten Daten-Sets oder aus einer Datei das Daten-Set aus, das mit der Arbeitsdatei zusammengefügt werden soll.
Ein geöffnetes Daten-Set
region2.sav [DataSet2]
Eine externe SPSS-Datendatei
Durchsuchen...
Datendateien, die nicht im SPSS-Format vorliegen, müssen erst in SPSS geöffnet werden, bevor sie beim Zusammenfügen verwendet werden können.
Weiter
Abbrechen
Hilfe

Fälle hinzufügen

Mithilfe von “Fälle hinzufügen” wird die Arbeitsdatei mit einem zweiten Daten-Set oder einer Datendatei im SPSS-Format zusammengefügt, das bzw. die dieselben Variablen (Spalten) aber unterschiedliche Fälle (Zeilen) enthält. So könnten Sie zum Beispiel dieselben Informationen für Kunden in zwei verschiedenen Verkaufsgebieten aufzeichnen und die Daten für jedes Gebiet in getrennten Dateien speichern. Das zweite Daten-Set kann eine externe Datendatei im SPSS-Format oder ein Daten-Set aus der aktuellen Sitzung sein.

Abbildung 9-3
Dialogfeld "Fälle hinzufügen"



Nicht gepaarte Variablen. Diese Variablen werden nicht in die neue, zusammengefügte Datendatei aufgenommen. Die Variablen aus der Arbeitsdatei sind mit einem Sternchen (*) gekennzeichnet. Die Variablen aus dem anderen Daten-Set sind mit einem Pluszeichen (+) gekennzeichnet. Diese Liste enthält in der Standardeinstellung folgende Einträge:

- Variablen, die nicht mit gleichem Namen in beiden Datendateien enthalten sind. Sie können Paare aus den nicht gepaarten Variablen bilden und diese in die neue, zusammengefügte Datei aufnehmen.
- Variablen, die als numerische Daten in der einen Datei und als String-Daten in der anderen Datei definiert sind. Numerische Variablen können nicht mit String-Variablen zusammengefügt werden.
- String-Variablen ungleicher Länge. Die definierte Länge einer String-Variablen muß in beiden Datendateien gleich sein.

Variablen in neuer Arbeitsdatei. Diese Variablen werden in die neue, zusammengefügte Datendatei aufgenommen. Die Standardeinstellung sieht vor, daß alle Variablen in die Liste aufgenommen werden, die sowohl in ihrem Namen als auch im Datentyp (numerisch oder String) übereinstimmen.

- Sie können Variablen, die nicht in die zusammengeführte Datei aufgenommen werden sollen, aus der Liste entfernen.
- Alle nicht gepaarten Variablen, die in die zusammengefügte Datei aufgenommen werden, enthalten fehlende Daten für die Fälle aus der Datei, die diese Variable nicht enthält.

Datei-Indikator als Variable. Gibt für jeden Fall die Quelldatei an. Diese Variable hat den Wert 0 für Fälle aus der Arbeitsdatei, und den Wert 1 für Fälle aus der externen Datendatei.

So fügen Sie Datendateien mit denselben Variablen und unterschiedlichen Fällen zusammen

▶ Öffnen Sie mindestens eine der Datendateien, die Sie zusammenfügen möchten. Wenn Sie mehrere Daten-Sets geöffnet haben, legen Sie eines der zusammenzufügenden Daten-Sets als Arbeitsdatei fest. Die Fälle aus dieser Datei werden zuerst in die neue, zusammengefügte Datendatei übernommen.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Daten
  Dateien zusammenfügen
    Fälle hinzufügen

▶ Wählen Sie das Daten-Set oder die Datendatei im SPSS-Format aus, das bzw. die Sie mit der Arbeitsdatei zusammenfügen möchten.

▶ Entfernen Sie alle Variablen, die nicht übernommen werden sollen, aus der Liste "Variablen in neuer Arbeitsdatei".

▶ Fügen Sie gegebenenfalls die Variablenpaare aus der Liste "Nicht gepaarte Variablen" hinzu, die dieselben Informationen unter verschiedenen Variablennamen in den beiden Dateien enthalten. So könnte zum Beispiel das Geburtsdatum in der einen Datei den Variablennamen *Gebdat* haben und in der anderen Datei den Namen *DatGeb*.

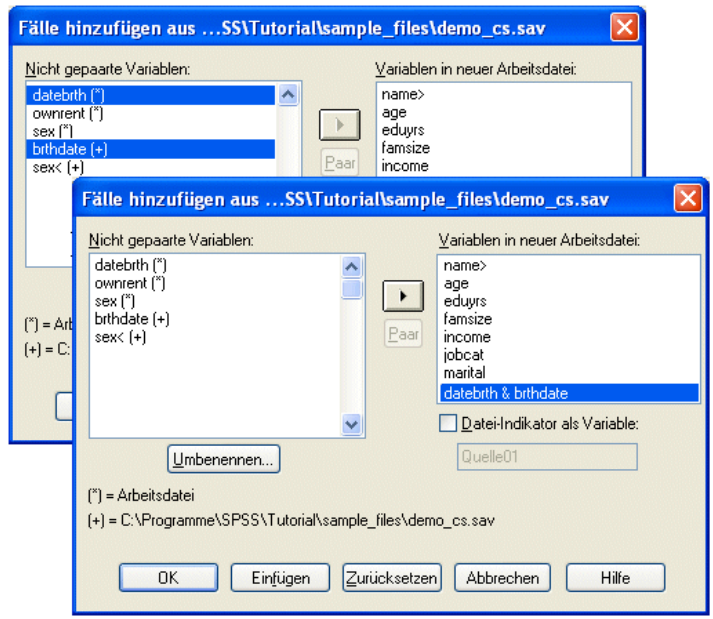
So wählen Sie ein Paar nicht gepaarter Variablen aus

▶ Klicken Sie in der Liste "Nicht gepaarte Variablen" auf eine Variable.

▶ Klicken Sie bei gedrückter STRG-Taste auf die andere Variable in der Liste. (Drücken Sie dazu gleichzeitig auf die STRG-Taste und die linke Maustaste.)

- Klicken Sie auf Paar, um das Variablenpaar in die Liste “Variablen in neuer Arbeitsdatei” zu verschieben. (Der Variablenname aus der Arbeitsdatei wird in der zusammengefügten Datei als Variablenname verwendet.)

Abbildung 9-4
Auswählen von Variablenpaaren durch Klicken mit der Maust bei gedrückter STRG-Taste



Fälle hinzufügen: Umbenennen

Sie können Variablen sowohl in der Arbeitsdatei als auch im anderen Daten-Set umbenennen, bevor Sie diese aus der Liste der nicht gepaarten Variablen in die Liste der Variablen verschieben, die in die zusammengeführte Datei aufgenommen werden. Mit dem Umbenennen von Variablen erreichen Sie folgendes:

- Sie können die Variablennamen aus dem anderen Daten-Set anstelle der Namen aus der Arbeitsdatei für Variablenpaare verwenden.
- Sie können zwei Variablen mit demselben Namen, aber unterschiedlichen Typen oder verschiedenen String-Längen aufnehmen. Wenn Sie zum Beispiel sowohl die numerische Variable *Geschl* aus der Arbeitsdatei als auch die String-Variable *Geschl* aus dem anderen Daten-Set aufnehmen möchten, müssen Sie zuerst eine der beiden Variablen umbenennen.

Fälle hinzufügen: Informationen aus dem Datenlexikon

Alle vorhandenen Informationen aus dem Datenlexikon der Arbeitsdatei (Variablen- und Wertelabels, benutzerdefinierte fehlende Werte und Anzeigenformate) werden der zusammengefügten Datendatei zugewiesen.

- Wenn für eine Variable keine Informationen im Datenlexikon der Arbeitsdatei definiert sind, werden die entsprechenden Informationen aus dem Datenlexikon des anderen Daten-Sets verwendet.
- Wenn die Arbeitsdatei definierte Wertelabels oder benutzerdefinierte fehlende Werte für eine Variable enthält, werden alle weiteren Wertelabels oder benutzerdefinierten fehlenden Werte für diese Variable im anderen Daten-Set ignoriert.

Zusammenfügen von mehr als zwei Datenquellen

Mithilfe der Befehlssyntax können Sie bis zu 50 Daten-Sets und/oder Datendateien zusammenfügen. Weitere Informationen finden Sie unter dem Befehl `ADD FILES` in der *SPSS Command Syntax Reference* (verfügbar über das Menü “Hilfe”).

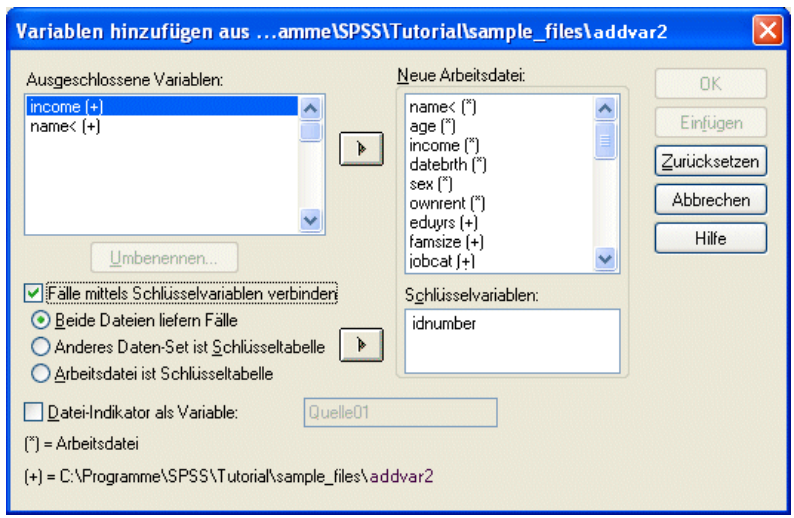
Variablen hinzufügen

Mithilfe von “Variablen hinzufügen” wird die Arbeitsdatei mit einem weiteren geöffneten Daten-Set oder einer Datendatei im SPSS-Format zusammengefügt, das bzw. die dieselben Fälle (Zeilen) aber verschiedene Variablen (Spalten) enthält. So könnten Sie zum Beispiel eine Datendatei mit Ergebnissen vor einem Test mit einer Datei zusammenfügen, welche die Ergebnisse nach einem Test enthält.

- Die Fälle müssen in beiden Daten-Sets in derselben Reihenfolge sortiert sein.
- Wenn eine oder mehrere Schlüsselvariablen zum Zuordnen der Fälle verwendet werden, müssen die beiden Daten-Sets in aufsteigender Reihenfolge der Schlüsselvariable(n) sortiert sein.
- Variablennamen in der zweiten Datendatei, die mit Variablennamen in der Arbeitsdatei übereinstimmen, werden in der Standardeinstellung ausgeschlossen, da bei “Variablen hinzufügen” davon ausgegangen wird, dass diese Variablen dieselben Informationen enthalten.

Datei-Indikator als Variable. Gibt für jeden Fall die Quelldatei an. Diese Variable hat den Wert 0 für Fälle aus der Arbeitsdatei, und den Wert 1 für Fälle aus der externen Datendatei.

Abbildung 9-5
Dialogfeld “Variablen hinzufügen”



Ausgeschlossene Variablen. Diese Variablen werden nicht in die neue, zusammengefügte Datendatei aufgenommen. Diese Liste enthält in der Standardeinstellung alle Variablennamen aus dem anderen Daten-Set, die mit Variablennamen aus der Arbeitsdatei übereinstimmen. Die Variablen aus der Arbeitsdatei sind mit einem Sternchen (*) gekennzeichnet. Die Variablen aus dem anderen Daten-Set sind mit einem Pluszeichen (+) gekennzeichnet. Wenn Sie eine ausgeschlossene Variable mit einem doppelt belegten Namen in die zusammengefügte Datei aufnehmen möchten, können Sie diese Variable umbenennen und sie der Liste einzuschließender Variablen hinzufügen.

Neue Arbeitsdatei. Diese Variablen werden in das neue, zusammengefügte Daten-Set aufgenommen. In der Standardeinstellung werden alle eindeutigen Variablennamen in beiden Daten-Sets in die Liste aufgenommen.

Schlüsselvariablen. Wenn einige Fälle in einem Daten-Set keine Entsprechung im anderen Daten-Set aufweisen (das heißt, einige Fälle fehlen in einem Daten-Set), verwenden Sie Schlüsselvariablen zum Identifizieren und richtigen Zuordnen der Fälle in den beiden Daten-Sets. Schlüsselvariablen können Sie außerdem bei Tabellenindexdateien verwenden.

- Die Schlüsselvariablen müssen in beiden Daten-Sets unter demselben Namen aufgeführt sein.
- Beide Daten-Sets müssen in aufsteigender Reihenfolge der Schlüsselvariablen sortiert sein, wobei die Variablen in der Liste “Schlüsselvariablen” entsprechend der Sortierfolge geordnet sein müssen.
- Fälle, die nicht in den Schlüsselvariablen übereinstimmen, werden in die zusammengefügte Datei aufgenommen, aber sie werden nicht mit Fällen aus der anderen Datei verknüpft. Fälle ohne Entsprechung enthalten nur Werte für die Variablen in der Datei, aus der sie stammen. Die Variablen aus der anderen Datei enthalten den systemdefinierten fehlenden Wert.

Anderes Daten-Set oder Arbeitsdatei ist Schlüsseltabelle. Eine Schlüsseltabelle, oder **Tabellenindexdatei**, ist eine Datei, deren Daten für jeden Fall mehreren Fällen der anderen Datendatei zugeordnet werden können. Wenn eine Datei zum Beispiel Informationen über einzelne Familienmitglieder (beispielsweise Geschlecht, Alter und Bildungsstand) und die andere Datei übergreifende Informationen zur ganzen Familie enthält (beispielsweise das Gesamteinkommen, die Familiengröße und den Wohnort), können Sie die Datei mit den Familiendaten als Tabellenindexdatei

verwenden. In der zusammengefügten Datendatei können Sie dann jedem einzelnen Familienmitglied die gemeinsamen Familiendaten zuweisen.

So fügen Sie Dateien mit gleichen Fällen, aber unterschiedlichen Variablen zusammen

- Öffnen Sie mindestens eine der Datendateien, die Sie zusammenfügen möchten. Wenn Sie mehrere Daten-Sets geöffnet haben, legen Sie eines der zusammenzufügenden Daten-Sets als Arbeitsdatei fest.
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Daten
  Dateien zusammenfügen
  Variablen hinzufügen...
- Wählen Sie das Daten-Set oder die Datendatei im SPSS-Format aus, das bzw. die Sie mit der Arbeitsdatei zusammenfügen möchten.

So wählen Sie Schlüsselvariablen aus

- Wählen Sie aus der Liste “Ausgeschlossene Variablen” die Variablen der externen Datei (mit “+” markiert) aus.
- Aktivieren Sie das Optionsfeld Fälle mittels Schlüsselvariablen verbinden.
- Fügen Sie die Variablen der Liste “Schlüsselvariablen” hinzu.

  Die Schlüsselvariablen müssen sowohl in der Arbeitsdatei als auch im anderen Daten-Set vorhanden sein. Beide Daten-Sets müssen in aufsteigender Reihenfolge der Schlüsselvariablen sortiert sein, wobei die Variablen in der Liste “Schlüsselvariablen” entsprechend der Sortierfolge geordnet sein müssen.

Variablen hinzufügen: Umbenennen

Sie können Variablen sowohl in der Arbeitsdatei als auch im anderen Daten-Set umbenennen, bevor Sie diese in die Liste der Variablen verschieben, die in die zusammengeführte Datei aufgenommen werden. Dies empfiehlt sich, wenn Sie zwei Variablen mit gleichem Namen aufnehmen möchten, die verschiedene Informationen in den beiden Dateien enthalten.

Zusammenfügen von mehr als zwei Datenquellen

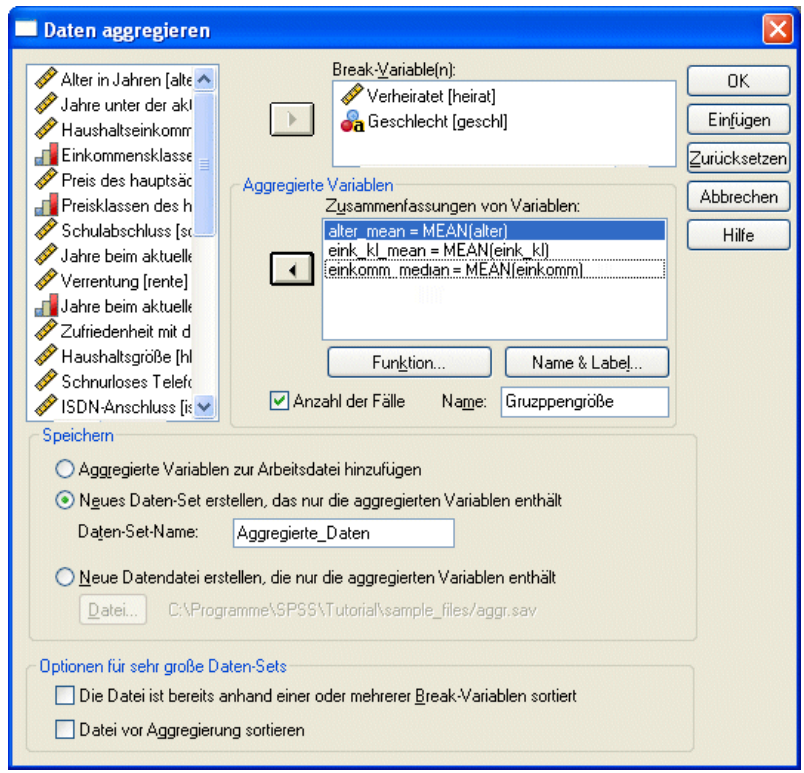
Mithilfe der Befehlssyntax können Sie bis zu 50 Daten-Sets und/oder Datendateien zusammenfügen. Weitere Informationen finden Sie unter dem Befehl `MATCH FILES` in der *SPSS Command Syntax Reference* (verfügbar über das Menü "Hilfe").

Daten aggregieren

Mit "Daten aggregieren" werden Fallgruppen in der Arbeitsdatei zu einzelnen Fällen kombiniert; hierbei wird eine neue, aggregierte Datei angelegt, oder es werden neue Variablen in der Arbeitsdatei angelegt, die aggregrierte Daten enthalten. Die Fälle werden nach einem oder mehreren Werten von Break-Variablen (Gruppenvariablen) aggregiert.

- Wenn Sie eine neue, aggregierte Datendatei anlegen, enthält diese neue Datei je einen Fall für jede Gruppe, die in den Break-Variablen definiert sind. Liegt beispielsweise eine Break-Variable mit zwei Werten vor, enthält die neue Datendatei nur zwei Fälle.
- Wenn Sie Aggregierungsvariablen in die Arbeitsdatei aufnehmen, wird die Datendatei selbst nicht aggregiert. Jeder Fall mit denselben Werten für die Break-Variable(n) erhält dieselben Werte für die neuen Aggregierungsvariablen. Wenn beispielsweise nur eine Break-Variable für das Geschlecht vorliegt, erhalten alle männlichen Personen denselben Wert für eine neue Aggregierungsvariable, die das Durchschnittsalter erfasst.

Abbildung 9-6
Dialogfeld "Daten aggregieren"



Break-Variable(n). Die Fälle werden auf der Basis der Break-Variablen gruppiert. Jede eindeutige Kombination von Break-Variablenwerten definiert eine Gruppe. Wenn Sie eine neue, aggregierte Datendatei erstellen, werden alle Break-Variablen in der neuen Datei unter dem bisherigen Namen und mit den vorhandenen Informationen aus dem Datenlexikon gespeichert. Die Break-Variablen können numerische Variablen oder String-Variablen sein.

Variablen aggregieren. Quellvariablen werden in Verbindung mit Aggregierungsfunktionen zum Erzeugen der neuen Aggregierungsvariablen herangezogen. Der Name der Aggregierungsvariablen wird von einem optionalen Variablenlabel in Anführungszeichen, dem Namen der Aggregierungsfunktion und der Quellvariablen in Klammern gefolgt. Quellvariablen für Aggregierungsfunktionen müssen numerisch sein.

Sie können die vorgegebenen Namen für die Aggregierungsvariablen mit neuen Variablennamen überschreiben, aussagekräftige Variablenlabels verwenden und die Funktionen zum Berechnen der aggregierten Datenwerte ändern. Sie können auch eine Variable anlegen, welche die Anzahl der Fälle in jeder Break-Gruppe enthält.

So aggregieren Sie eine Datendatei:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Daten
  Aggregieren...

- Wählen Sie eine oder mehrere Break-Variablen aus, die definieren, wie die Fälle zum Erzeugen der aggregierten Daten gruppiert werden.

- Wählen Sie mindestens eine Aggregierungsvariable aus.

- Wählen Sie für jede Aggregierungsvariable eine Aggregierungsfunktion aus.

Speichern von aggregierten Ergebnissen

Sie können wahlweise Aggregierungsvariablen in die Arbeitsdatei aufnehmen oder eine neue, aggregierte Datendatei erzeugen.

- **Aggregierte Variablen der Arbeitsdatei hinzufügen.** Neue Variablen, die auf Aggregierungsfunktionen beruhen, werden der Arbeitsdatei hinzugefügt. Die Datendatei selbst wird nicht aggregiert. Jeder Fall mit demselben Wert bzw. denselben Werten der Break-Variablen erhält die gleichen Werte für die neuen Aggregatvariablen.
- **Neues Daten-Set erstellen, das nur die aggregierten Variablen enthält.** Speichert die aggregierten Daten in ein neues Datenblatt in der aktuellen Sitzung. Das Datenblatt enthält die Break-Variablen, die die aggregierten Fälle bestimmen und alle aggregierten Variablen, die durch Aggregierungsfunktionen festgelegt werden. Das aktive Datenblatt bleibt davon unberührt.
- **Neue Datendatei erstellen, die nur die aggregierten Variablen enthält.** Erstellt eine neue Datendatei mit den aggregierten Daten. In der Standardeinstellung wird eine Datei mit dem Namen "aggr.sav" im aktuellen Verzeichnis gespeichert.

Sortieroptionen für umfangreiche Datendateien

Bei äußerst umfangreichen Datendateien empfiehlt es sich, die Dateien vor der Aggregierung zu sortieren.

Datei ist bereits nach Break-Variablen sortiert. Wenn die Daten bereits nach den Werten der Break-Variablen sortiert wurden, sorgt diese Option für einen schnelleren Ablauf der Prozedur und geringeren Speicherplatzbedarf. Diese Option sollte mit Bedacht verwendet werden.

- Die Daten müssen nach den Werten der Break-Variablen sortiert werden, und zwar in derselben Reihenfolge wie die Break-Variablen, die für die Funktion "Daten aggregieren" angegeben wurden.
- Wenn Sie Variablen in die Arbeitsdatei aufnehmen, wählen Sie diese Option nur dann aus, wenn die Daten anhand der Werte der Break-Variablen in aufsteigender Reihenfolge sortiert sind.

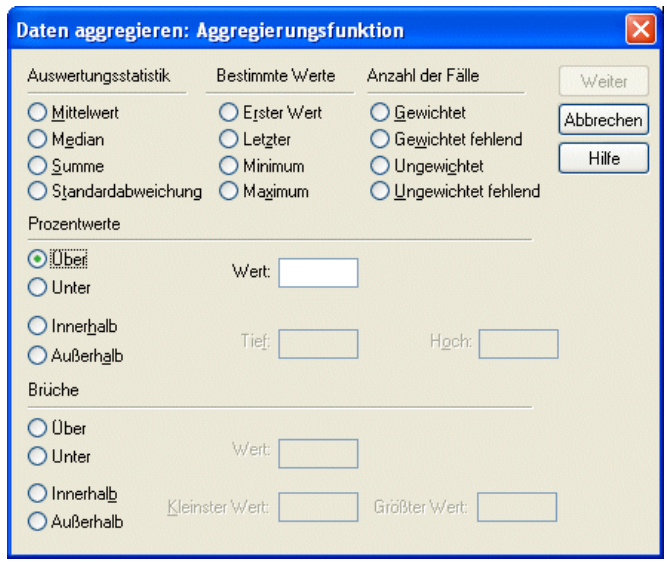
Datei vor dem Aggregieren sortieren. In seltenen Fällen kann es bei großen Datendateien nötig sein, die Datendatei vor dem Aggregieren nach den Werten der Break-Variablen zu sortieren. Diese Option ist nur dann angeraten, wenn Speicherplatz- bzw. Leistungsprobleme auftreten.

Daten aggregieren: Aggregierungsfunktion

In diesem Dialogfeld legen Sie die Funktion fest, die zum Berechnen der aggregierten Datenwerte für die Variablen verwendet wird, die in der Liste "Variablen aggregieren" aus dem Dialogfeld "Daten aggregieren" ausgewählt wurden. Folgende Aggregierungsfunktionen stehen zur Verfügung:

- Auswertungsfunktionen, einschließlich Mittelwert, Median, Standardabweichung und Summe.
- Anzahl der Fälle, einschließlich ungewichtet, gewichtet, nichtfehlend und fehlend.
- Prozentsatz oder Anteil von Werten über oder unter einem festgelegten Wert.
- Prozentsatz oder Anteil von Werten innerhalb oder außerhalb eines festgelegten Bereichs.

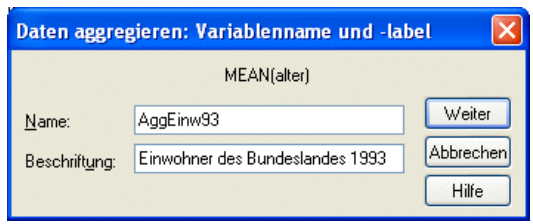
Abbildung 9-7
Dialogfeld "Daten aggregieren: Aggregierungsfunktion"



Daten aggregieren: Variablenname und -label

Mit "Daten aggregieren" werden den aggregierten Variablen in der neuen Datendatei vorgegebene Variablennamen zugewiesen. In diesem Dialogfeld können Sie den Variablennamen der ausgewählten Variablen in der Liste "Variablen aggregieren" ändern und ein aussagekräftiges Variablenlabel angeben. Für weitere Informationen siehe "Variablennamen" in Kapitel 5 auf S. 99.

Abbildung 9-8
Dialogfeld "Daten aggregieren: Variablenname und -label"

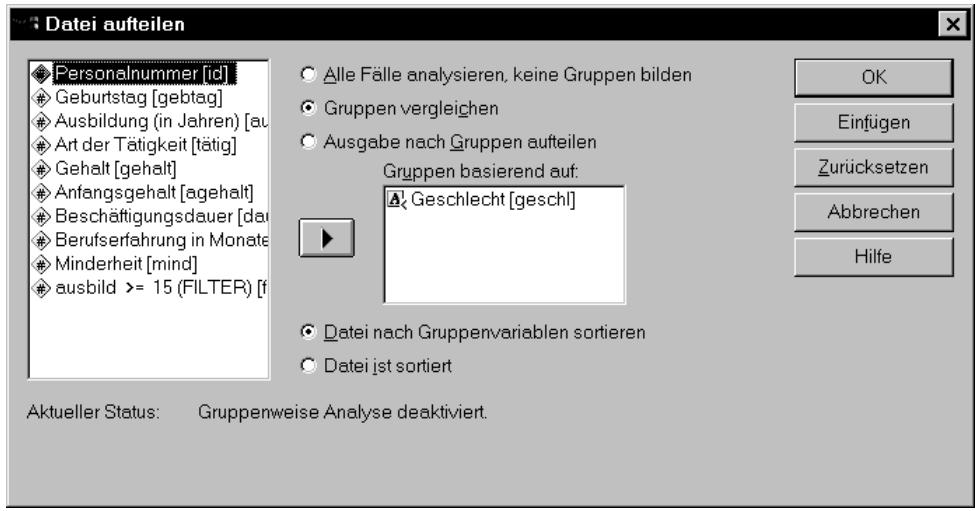


Datei aufteilen

Mit “Datei aufteilen” wird die Datendatei in separate Gruppen für die Analyse aufgeteilt, und zwar nach dem Wert einer oder mehrerer Gruppenvariablen. Wenn Sie mehrere Gruppenvariablen auswählen, werden die Fälle nach jeder Variablen innerhalb von Kategorien der vorhergehenden Variablen in der Liste “Gruppen basierend auf” geordnet. Wenn Sie zum Beispiel *Geschl* als erste Sortiervariable auswählen und *Minder* als zweite, wird beim Sortieren innerhalb der Geschlechtskategorien nach Minderheit sortiert.

- Sie können bis zu acht Gruppenvariablen festlegen.
- Bei langen String-Variablen (String-Variablen mit mehr als 8 Zeichen) gelten jeweils acht Zeichen als eine Variable, bis zur Obergrenze von acht Gruppenvariablen.
- Die Fälle müssen nach den Werten der Gruppenvariablen und in derselben Reihenfolge sortiert werden wie die in der Liste “Gruppen basierend auf” aufgeführten Variablen. Wenn die Datendatei noch nicht sortiert ist, klicken Sie auf Datei nach Gruppenvariablen sortieren.

Abbildung 9-9
Dialogfeld “Datei aufteilen”



Gruppen vergleichen. Die Gruppen der aufgeteilten Datei werden zu Vergleichszwecken zusammen angezeigt. Bei Pivot-Tabellen wird eine einzelne Pivot-Tabelle angelegt. Die Variablen zum Aufteilen der Datei können zwischen den Tabellendimensionen verschoben werden. Bei Diagrammen wird zu jeder Gruppe der aufgeteilten Datei ein separates Diagramm erstellt. Die Diagramme werden zusammen im Viewer angezeigt.

Ausgabe nach Gruppen aufteilen. Die Ergebnisse jeder Prozedur werden für jede Gruppe einer aufgeteilten Datei separat angezeigt.

So teilen Sie eine Datendatei für die Analyse auf:

- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Daten
  Datei aufteilen

- Aktivieren Sie das Optionsfeld Gruppen vergleichen oder Ausgabe nach Gruppen aufteilen.

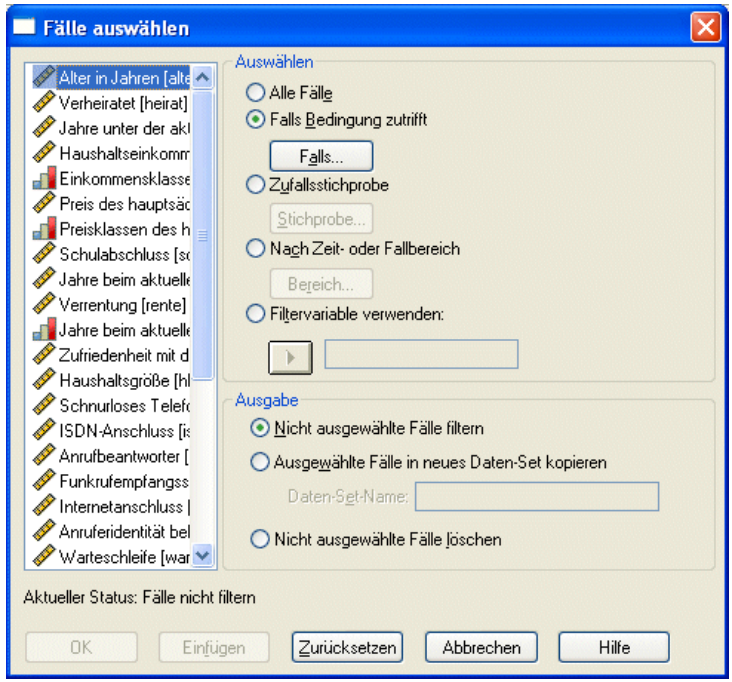
- Wählen Sie eine oder mehrere Gruppenvariablen aus.

Fälle auswählen

Im Dialogfeld “Fälle auswählen” stehen Ihnen mehrere Methoden zum Auswählen einer Untergruppe von Fällen zur Verfügung. Diese Methoden basieren auf Kriterien, die unter anderem Variablen und komplexe Ausdrücke zulassen. Sie können auch eine Zufallsstichprobe aus den Fällen auswählen. Die Kriterien zum Festlegen der Untergruppen können folgende Elemente enthalten:

- Variablenwerte und -bereiche
- Datums- und Zeitbereiche
- Fallnummern (Zeilennummern)
- Arithmetische Ausdrücke
- Logische Ausdrücke
- Funktionen

Abbildung 9-10
Dialogfeld "Fälle auswählen"



Alle Fälle. Deaktiviert die Filterung der Fälle und verwendet alle Fälle.

Falls Bedingung zutrifft. Verwendet eine Bedingung zur Auswahl von Fällen. Wenn das Ergebnis der Bedingung wahr ist, wird der Fall ausgewählt. Wenn das Ergebnis falsch oder fehlend ist, wird der Fall nicht ausgewählt.

Zufallsstichprobe. Auswahl einer Zufallsstichprobe, deren Fallzahl näherungsweise durch einen Prozentsatz oder durch die genaue Anzahl vorgegeben werden kann.

Nach Zeit- oder Fallbereich. Wählt Fälle auf der Grundlage eines Bereichs von Fallnummern oder eines Datums- bzw. Zeitbereichs aus.

Filtervariable verwenden. Verwendet die ausgewählte, numerische Variable aus der Datendatei als Filtervariable. Fälle mit einem anderen Wert als 0 oder fehlend, werden ausgewählt.

Ausgabe

In diesem Abschnitt wird die Behandlung von nicht ausgewählten Fällen festgelegt. Die folgenden Optionen stehen für die Behandlung nicht ausgewählter Fälle zur Auswahl:

- **Filtern von nicht ausgewählten Fällen.** Nicht ausgewählte Fälle werden nicht in die Analyse aufgenommen, verbleiben jedoch im Daten-Set. Sie können die nicht ausgewählten Fälle später in der Sitzung verwenden, wenn Sie die Filterfunktion deaktivieren. Wenn Sie eine Zufallsstichprobe oder Fälle anhand eines bedingten Ausdrucks auswählen, wird die Variable *filter_$* mit dem Wert 1 für ausgewählte Fälle und dem Wert 0 für nicht ausgewählte Fälle erzeugt.
- **Kopieren der ausgewählten Fälle in ein neues Daten-Set.** Die ausgewählten Fälle werden in ein neues Daten-Set kopiert, und das ursprüngliche Daten-Set verbleibt unverändert. Nicht ausgewählte Fälle werden nicht in das neue Daten-Set aufgenommen. Sie verbleiben im ursprünglichen Zustand im ursprünglichen Daten-Set.
- **Löschen von nicht ausgewählten Fällen.** Nicht ausgewählte Fälle werden aus dem Daten-Set gelöscht. Gelöschte Fälle können nur wiederhergestellt werden, indem Sie die Datei ohne Speichern der Änderungen schließen und sie dann erneut öffnen. Wenn Sie die Änderungen in der Datendatei speichern, werden die Fälle dauerhaft gelöscht. *Anmerkung*: Wenn Sie nicht ausgewählte Fälle löschen und die Datei speichern, können die Fälle nicht wiederhergestellt werden.

So wählen Sie eine Teilmenge von Fällen aus:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Daten
  Fälle auswählen

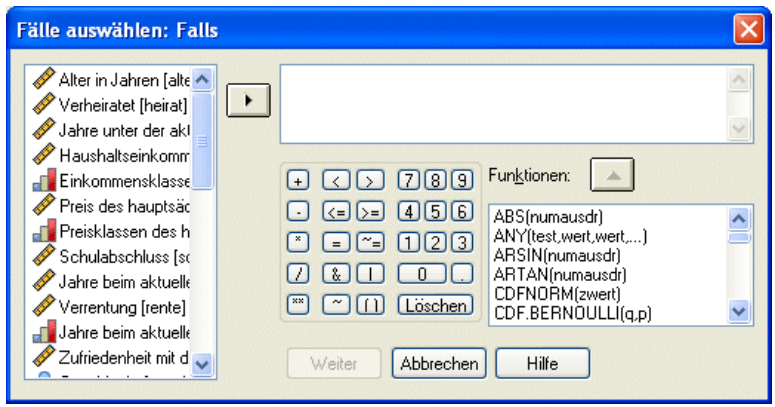
▶ Wählen Sie eine der Methoden zum Auswählen von Fällen.

▶ Geben Sie die Kriterien für die Auswahl der Fälle an.

Fälle auswählen: Falls

In diesem Dialogfeld können Sie anhand eines bedingten Ausdrucks eine Teilmenge der Fälle auswählen. Ein bedingter Ausdruck gibt für jeden Fall den Wert *Wahr*, *Falsch* oder *Fehlend* zurück.

Abbildung 9-11
Dialogfeld "Fälle auswählen: Falls"

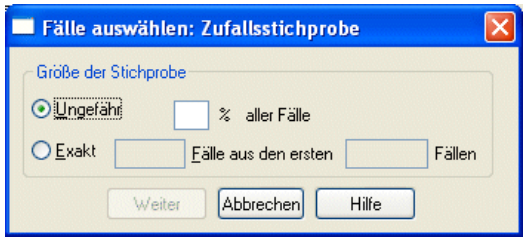


- Wenn das Ergebnis eines bedingten Ausdrucks *Wahr* ist, wird der Fall in die ausgewählte Untergruppe aufgenommen.
- Wenn das Ergebnis eines bedingten Ausdrucks *Falsch* oder *Fehlend* ist, wird der Fall nicht in die ausgewählte Untergruppe aufgenommen.
- In den meisten bedingten Ausdrücke wird mindestens einer der sechs Vergleichsoperatoren (<, >, <=, >=, = und ~=) verwendet. Diese sind auf der Rechentastatur verfügbar.
- Bedingte Ausdrücke können Variablennamen, Konstanten, arithmetische Operatoren, numerische (und andere) Funktionen, logische Variablen und Vergleichsoperatoren enthalten.

Fälle auswählen: Zufallsstichprobe

In diesem Dialogfeld können Sie eine Zufallsstichprobe nach einem ungefähren Prozentsatz oder einer genauen Anzahl von Fällen auswählen. Für die Stichproben erfolgt keine Ersetzung, so daß ein Fall jeweils nur einmal ausgewählt werden kann.

Abbildung 9-12
Dialogfeld "Fälle auswählen: Zufallsstichprobe"



Ungefähr. Erstellt eine Zufallsstichprobe, die ungefähr den angegebenen Prozentsatz aller Fälle enthält. Da SPSS für jeden Fall eine unabhängige Pseudo-Zufallsentscheidung trifft, entspricht der Prozentsatz der tatsächlich ausgewählten Fälle nur ungefähr dem angegebenen Prozentwert. Je mehr Fälle sich in der Datendatei befinden, desto eher entspricht der Prozentsatz ausgewählter Fälle dem angegebenen Prozentsatz.

Exakt. Geben Sie die gewünschte Anzahl der Fälle ein. Sie müssen außerdem die Anzahl der Fälle angeben, aus denen die Stichprobe gezogen werden soll. Diese zweite Zahl muß kleiner oder gleich der Gesamtanzahl der Fälle in der Datendatei sein. Wenn die angegebene Anzahl die Gesamtanzahl der Fälle in der Datendatei übersteigt, enthält die Stichprobe entsprechend weniger Fälle als die geforderte Anzahl.

Fälle auswählen: Bereich

In diesem Dialogfeld können Sie Fälle anhand eines Bereichs von Fallnummern oder eines Zeitbereichs auswählen.

- Die Fallbereiche basieren auf der Zeilennummer, die im Daten-Editor angezeigt wird.
- Datums- und Zeitbereiche sind nur für **Zeitreihendaten** mit definierten Datumsvariablen verfügbar (Menü "Daten", Befehl "Datum definieren").

Abbildung 9-13
Dialogfeld "Fälle auswählen: Bereich" für Fallbereich (keine definierten Datumsvariablen)

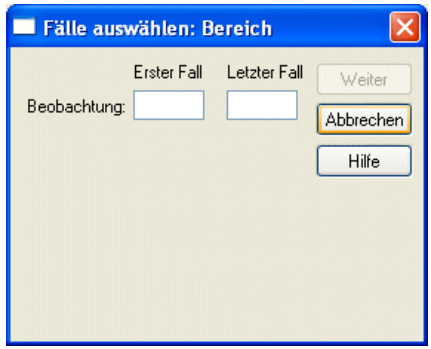


Abbildung 9-14
Dialogfeld "Fälle auswählen: Bereich" für Zeitreihendaten mit definierten Datumsvariablen

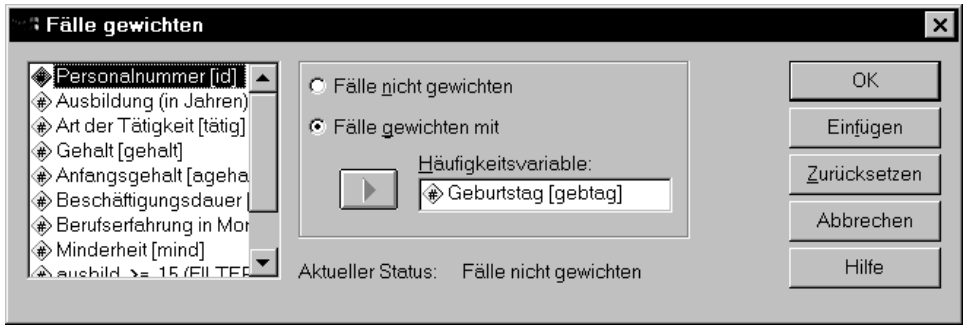
Fälle gewichten

Mit "Fälle gewichten" werden Fälle für die statistische Analyse unterschiedlich gewichtet (durch simulierte Replikation).

- Die Werte der Gewichtungsvariablen müssen der Anzahl der Beobachtungen entsprechen, die durch einzelne Fälle in der Datendatei dargestellt wird.

- Fälle, bei denen die Gewichtungsvariablen den Wert Null, einen negativen Wert oder einen fehlenden Wert aufweisen, werden von der Analyse ausgeschlossen.
- Gebrochene Zahlen sind zulässig. Sie werden nur verwendet, wenn dies sinnvoll ist, und höchstwahrscheinlich bei Fällen in Tabellen.

Abbildung 9-15
Dialogfeld "Fälle gewichten"



Sobald Sie eine Gewichtungsvariable zugewiesen haben, bleibt diese solange wirksam, bis Sie eine andere Gewichtungsvariable auswählen oder die Gewichtung ausschalten. Wenn Sie eine gewichtete Datendatei speichern, werden die Informationen für die Gewichtung zusammen mit der Datendatei gespeichert. Sie können die Gewichtung jederzeit aufheben, selbst wenn die Datei bereits in gewichteter Form gespeichert wurde.

Gewichtungen in der Prozedur "Kreuztabellen". In der Kreuztabellen-Prozedur gibt es mehrere Optionen für die Behandlung von Fallgewichtungen. Für weitere Informationen siehe "Kreuztabellen: Zellenanzeige" in Kapitel 17 auf S. 395.

Gewichtungen bei Streudiagrammen und Histogrammen. Bei Streudiagrammen und Histogrammen gibt es eine Option zum Ein- und Ausschalten von Fallgewichtungen. Dies betrifft jedoch nicht die Fälle, bei denen die Gewichtungsvariable den Wert Null oder einen negativen oder fehlenden Wert aufweist. Diese Fälle bleiben selbst dann vom Diagramm ausgeschlossen, wenn Sie die Gewichtung aus dem Diagramm heraus ausschalten.

So gewichten Sie Fälle:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Daten
  Fälle gewichten...

- Aktivieren Sie das Optionsfeld Fälle gewichten mit.
- Wählen Sie eine Häufigkeitsvariable aus.

  Die Werte der Häufigkeitsvariablen werden als Fallgewichtungen verwendet. So steht zum Beispiel ein Fall mit dem Wert 3 bei der Häufigkeitsvariablen in der gewichteten Datendatei für drei Fälle.

Umstrukturieren von Daten

Verwenden Sie den Assistenten für die Datenumstrukturierung, um Daten so umzustrukturieren, daß sie in der gewünschten Prozedur von SPSS verwendet werden können. Der Assistent ersetzt die aktuelle Datei durch eine neue, umstrukturierte Datei. Der Assistent stellt Ihnen folgende Optionen zur Verfügung:

- Umstrukturieren ausgewählter Variablen in Fälle.
- Umstrukturieren ausgewählter Fälle in Variablen.
- Transponieren sämtlicher Daten.

So strukturieren Sie Daten um

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Daten
  Umstrukturieren...
- Wählen Sie den Typ der Umstrukturierung aus, den Sie durchführen möchten.
- Wählen Sie die Daten aus, die umstrukturiert werden sollen.

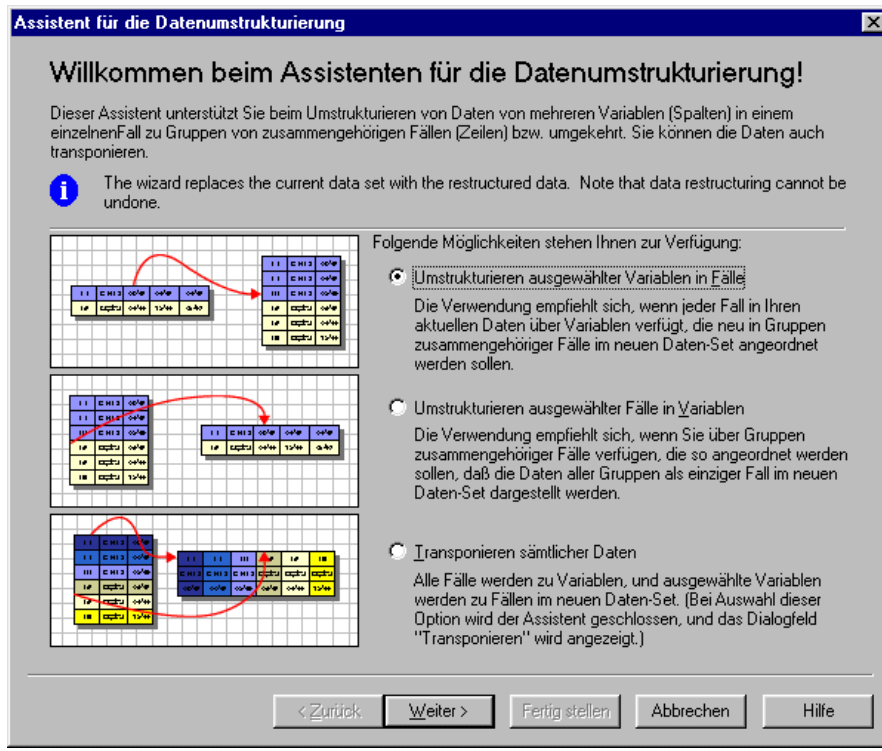
  Die folgenden Optionen sind verfügbar:

  - Erstellen von Bezeichnervariablen, mit deren Hilfe Sie einen Wert aus der neuen Datei zu einem Wert in der ursprünglichen Datei zurückverfolgen können.
  - Sortieren der Daten vor der Umstrukturierung.
  - Festlegen von Optionen für die neue Datei.
  - Einfügen der Befehlssyntax in ein Syntax-Fenster.

Assistent für die Datenumstrukturierung: Auswählen des Typs

Mit dem Assistenten für die Datenumstrukturierung können Sie Daten umstrukturieren. Wählen Sie im ersten Dialogfeld den gewünschten Umstrukturierungstyp aus.

Abbildung 9-16
Assistent für die Datenumstrukturierung:


Assistent für die Datenumstrukturierung
Willkommen beim Assistenten für die Datenumstrukturierung!
Dieser Assistent unterstützt Sie beim Umstrukturieren von Daten von mehreren Variablen (Spalten) in einem einzelnenFall zu Gruppen von zusammengehörigen Fällen (Zeilen) bzw. umgekehrt. Sie können die Daten auch transponieren.
The wizard replaces the current data set with the restructured data. Note that data restructuring cannot be undone.
Folgende Möglichkeiten stehen Ihnen zur Verfügung:
Umstrukturieren ausgewählter Variablen in Fälle
Die Verwendung empfiehlt sich, wenn jeder Fall in Ihren aktuellen Daten über Variablen verfügt, die neu in Gruppen zusammengehöriger Fälle im neuen Daten-Set angeordnet werden sollen.
Umstrukturieren ausgewählter Fälle in Variablen
Die Verwendung empfiehlt sich, wenn Sie über Gruppen zusammengehöriger Fälle verfügen, die so angeordnet werden sollen, daß die Daten aller Gruppen als einziger Fall im neuen Daten-Set dargestellt werden.
Transponieren sämtlicher Daten
Alle Fälle werden zu Variablen, und ausgewählte Variablen werden zu Fällen im neuen Daten-Set. (Bei Auswahl dieser Option wird der Assistent geschlossen, und das Dialogfeld "Transponieren" wird angezeigt.)
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

- **Umstrukturieren ausgewählter Variablen in Fälle.** Wählen Sie diese Option aus, wenn die Daten Gruppen verwandter Spalten enthalten, die in der neuen Datendatei als Gruppen von Zeilen dargestellt werden sollen. Bei Auswahl dieser Option zeigt der Assistent die Schritte für **Variablen zu Fällen** an.

- **Umstrukturieren ausgewählter Fälle in Variablen.** Wählen Sie diese Option aus, wenn die Daten Gruppen verwandter Zeilen enthalten, die in der neuen Datendatei als Gruppen von Spalten dargestellt werden sollen. Bei Auswahl dieser Option zeigt der Assistent die Schritte für **Fälle zu Variablen** an.
- **Transponieren sämtlicher Daten.** Wählen Sie diese Option, wenn Sie die Daten vertauschen möchten. Dadurch werden alle Zeilen in den neuen Daten zu Spalten und alle Spalten zu Zeilen. Bei Auswahl dieser Option wird der Assistent für die Datenumstrukturierung geschlossen und das Dialogfeld "Daten transponieren" geöffnet.

Festlegen der Art der Umstrukturierung der Daten

Eine **Variable** enthält zu analysierende Informationen, z. B. eine Messung oder einen Wert. Ein **Fall** ist eine Beobachtung, z. B. eine Person. In einer *einfachen* Datenstruktur ist jede Variable eine einzelne Spalte in den Daten, und jeder Fall bildet eine einzelne Zeile. Wenn Sie beispielsweise Testergebnisse für alle Schüler einer Klasse messen, werden alle Meßwerte in einer einzigen Spalte angezeigt, und jedem Schüler wird eine Zeile zugeordnet.

Beim Analysieren von Daten wird häufig untersucht, wie eine Variable in Abhängigkeit von einer bestimmten Bedingung variiert. Bei der Bedingung kann es sich um eine bestimmte zu erprobende Behandlung, eine demographische Gruppe, einen Zeitpunkt usw. handeln. In der Datenanalyse werden relevante Bedingungen häufig als **Faktoren** bezeichnet. Wenn Sie Faktoren analysieren, liegt eine *komplexe* Datenstruktur vor. Die Informationen zu einer Variablen können dabei in mehreren Datenspalten vorliegen (beispielsweise eine Spalte für jede Faktorstufe), oder es können in mehreren Zeilen Informationen zu einem Fall vorkommen (beispielsweise eine Zeile für jede Faktorstufe). Der Assistent für die Datenumstrukturierung unterstützt Sie bei der Umstrukturierung von Dateien mit einer komplexen Datenstruktur.

Welche Optionen Sie in diesem Assistenten auswählen, hängt von der Struktur der aktuellen Datei und der gewünschten Struktur der neuen Datei ab.

Wie sind die Daten in der aktuellen Datei angeordnet? Die aktuellen Daten können so angeordnet sein, daß Faktoren in einer *separaten* Variable (in Fallgruppen) oder *mit* der Variablen (in Variablengruppen) aufgezeichnet sind.

- **Fallgruppen.** Sind die Variablen und Bedingungen in der aktuellen Datei in verschiedenen Spalten aufgezeichnet? Ein Beispiel:

Variable	Faktor
8	1
9	1
3	2
1	2

In diesem Beispiel handelt es sich bei den ersten zwei Zeilen um eine **Fallgruppe**, da sie miteinander verbunden sind. Sie enthalten Daten für dieselbe Faktorstufe. Wenn die Daten in dieser Weise strukturiert sind, wird der Faktor in der Datenanalyse von SPSS oft als **Gruppenvariable** bezeichnet.

- **Spaltengruppen.** Sind die Variablen und Bedingungen der aktuellen Datei in derselben Spalte aufgezeichnet? Ein Beispiel:

var_1	var_2
8	3
9	1

In diesem Beispiel bilden die beiden Spalten eine **Variablengruppe**, da sie miteinander verbunden sind. Sie enthalten Daten für dieselbe Variable: *var_1* für Faktorstufe 1 und *var_2* für Faktorstufe 2. Wenn die Daten in dieser Weise strukturiert sind, wird der Faktor in der Datenanalyse von SPSS oft als **Meßwiederholung** bezeichnet.

Wie sollen die Daten in der neuen Datei angeordnet werden? Dies hängt in der Regel von der Prozedur ab, mit der Sie die Daten analysieren möchten.

- **Prozeduren, die Fallgruppen erfordern.** Zum Durchführen von Analysen, die eine Gruppenvariable erfordern, müssen die Daten in Fallgruppen angeordnet sein. Beispiele hierfür sind *univariate*, *multivariate* und *Varianzkomponenten* mit dem allgemeinen linearen Modell, gemischte Modelle und OLAP-Würfel sowie *unabhängige Stichproben* mit dem T-Test oder mit nichtparametrischen Tests.

Wenn Sie diese Analysen durchführen möchten und Ihre Daten in Form von Variablengruppen strukturiert sind, wählen Sie Umstrukturieren ausgewählter Variablen in Fälle aus.

- **Prozeduren, die Variablengruppen erfordern.** Zum Analysieren von Meßwiederholungen müssen die Daten in Variablengruppen angeordnet sein. Beispiele hierfür sind *Meßwiederholungen* mit dem allgemeinen linearen Modell, die Analyse von *zeitabhängigen Kovariaten* mit der Cox-Regressionsanalyse, *gepaarte Stichproben* mit dem T-Test oder *verbundene Stichproben* mit nichtparametrischen Tests. Wenn Sie diese Analysen durchführen möchten und Ihre Daten in Form von Fallgruppen strukturiert sind, wählen Sie Umstrukturieren ausgewählter Fälle in Variablen aus.

Beispiel für die Umstrukturierung von Variablen zu Fällen

In diesem Beispiel sind die Testergebnisse für jeden Faktor (*A* und *B*) in verschiedenen Spalten aufgezeichnet.

Abbildung 9-17
Aktuelle Daten für "Variablen zu Fälle"

	score_a	score_b
1	1014.00	864.00
2	684.00	636.00
3	810.00	638.00

Es soll ein *T*-Test bei unabhängigen Stichproben durchgeführt werden. Sie verfügen über eine Spaltengruppe, die aus *score_a* und *score_b* besteht, nicht jedoch über die für die Prozedur erforderliche **Gruppenvariable**. Wählen Sie im Assistenten für die Datenumstrukturierung Umstrukturieren ausgewählter Variablen in Fälle aus, strukturieren Sie eine Variablengruppe in eine neue Variable mit der Bezeichnung *score* um, und erstellen Sie einen Index unter dem Namen *group*. Die neue Datendatei wird in der folgenden Abbildung dargestellt.

Abbildung 9-18
Neue umstrukturierte Daten für "Variablen zu Fälle"

	group	score
1	SCORE_A	1014.00
2	SCORE_B	864.00
3	SCORE_A	684.00
4	SCORE_B	636.00
5	SCORE_A	810.00
6	SCORE_B	638.00

Wenn Sie den *T*-Test bei unabhängigen Stichproben ausführen, können Sie nun *group* als Gruppenvariable verwenden.

Beispiel für die Umstrukturierung von Fällen zu Variablen

In diesem Beispiel werden die Testergebnisse für jedes Subjekt zweimal aufgezeichnet, und zwar vor und nach einer Behandlung.

Abbildung 9-19
Aktuelle Daten für "Fälle zu Variablen"

	id	wert	zeit
1	1	1014.00	bef
2	1	864.00	aft
3	2	684.00	bef
4	2	636.00	aft

Es soll ein *T*-Test bei gepaarten Stichproben durchgeführt werden. Die Daten sind in Fallgruppen angeordnet, es fehlen jedoch die **Meßwiederholungen** für die Variablenpaare, die für die Prozedur erforderlich sind. Wählen Sie im Assistenten für die Datenumstrukturierung Umstrukturieren ausgewählter Fälle in Variablen aus, verwenden Sie *id* zum Identifizieren der Zeilengruppen in den aktuellen Daten und *time*, um die Variablengruppe in der neuen Datei zu erstellen.

Abbildung 9-20
Neue umstrukturierte Daten für "Fälle zu Variablen"

	id	wertnach	wertvorh
1	1	864.00	1014.00
2	2	636.00	684.00

Wenn Sie den *T*-Test bei gepaarten Stichproben ausführen, können Sie nun *Vorh* und *Nach* als Variablenpaar verwenden.

Assistent für die Datenumstrukturierung (Variablen zu Fällen): Anzahl von Variablengruppen

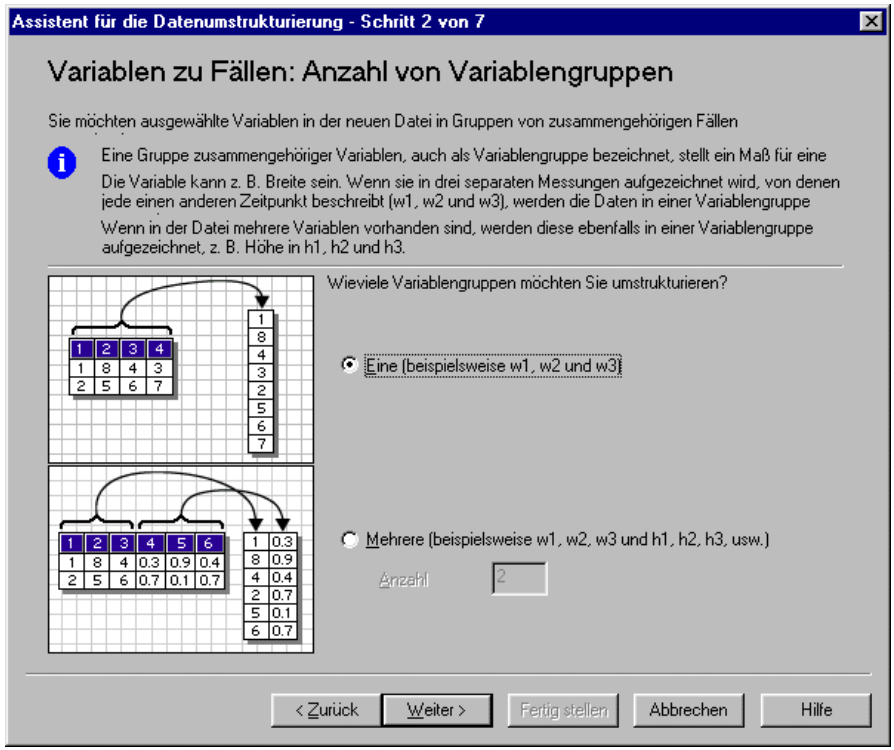
Anmerkung: Der Assistent zeigt diesen Schritt an, wenn Variablengruppen in Zeilen umstrukturiert werden sollen.

Legen Sie in diesem Schritt fest, wie viele Variablengruppen der aktuellen Datei in der neuen Datei umstrukturiert werden sollen.

Wie viele Variablengruppen gibt es in der aktuellen Datei? Stellen Sie fest, wie viele Variablengruppen in den aktuellen Daten vorliegen. Eine Gruppe verbundener Spalten, auch **Variablengruppe** genannt, zeichnet verschiedene Messungen derselben Variablen in mehreren Spalten auf. Wenn es in den aktuellen Daten beispielsweise drei Spalten mit den Bezeichnungen *w1*, *w2* und *w3* zum Aufzeichnen der **Breite** gibt, stellen diese Spalten eine Variablengruppe dar. Wenn drei weitere Spalten *h1*, *h2* und *h3* zum Aufzeichnen der **Höhe** vorhanden sind, verfügen Sie über zwei Variablengruppen.

Wie viele Variablengruppen soll die neue Datei enthalten?Legen Sie fest, wie viele Variablengruppen in der neuen Datendatei dargestellt werden sollen. Sie müssen nicht notwendigerweise alle Variablengruppen in die neue Datei umstrukturieren.

Abbildung 9-21
Assistent für die Datenumstrukturierung: Anzahl der Variablengruppen, Schritt 2


Assistent für die Datenumstrukturierung - Schritt 2 von 7
Variablen zu Fällen: Anzahl von Variablengruppen
Sie möchten ausgewählte Variablen in der neuen Datei in Gruppen von zusammengehörigen Fällen
Eine Gruppe zusammengehöriger Variablen, auch als Variablengruppe bezeichnet, stellt ein Maß für eine
Die Variable kann z. B. Breite sein. Wenn sie in drei separaten Messungen aufgezeichnet wird, von denen jede einen anderen Zeitpunkt beschreibt (w1, w2 und w3), werden die Daten in einer Variablengruppe
Wenn in der Datei mehrere Variablen vorhanden sind, werden diese ebenfalls in einer Variablengruppe aufgezeichnet, z. B. Höhe in h1, h2 und h3.
Wieviele Variablengruppen möchten Sie umstrukturieren?
Eine (beispielsweise w1, w2 und w3)
Mehrere (beispielsweise w1, w2, w3 und h1, h2, h3, usw.)
Anzahl
2
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

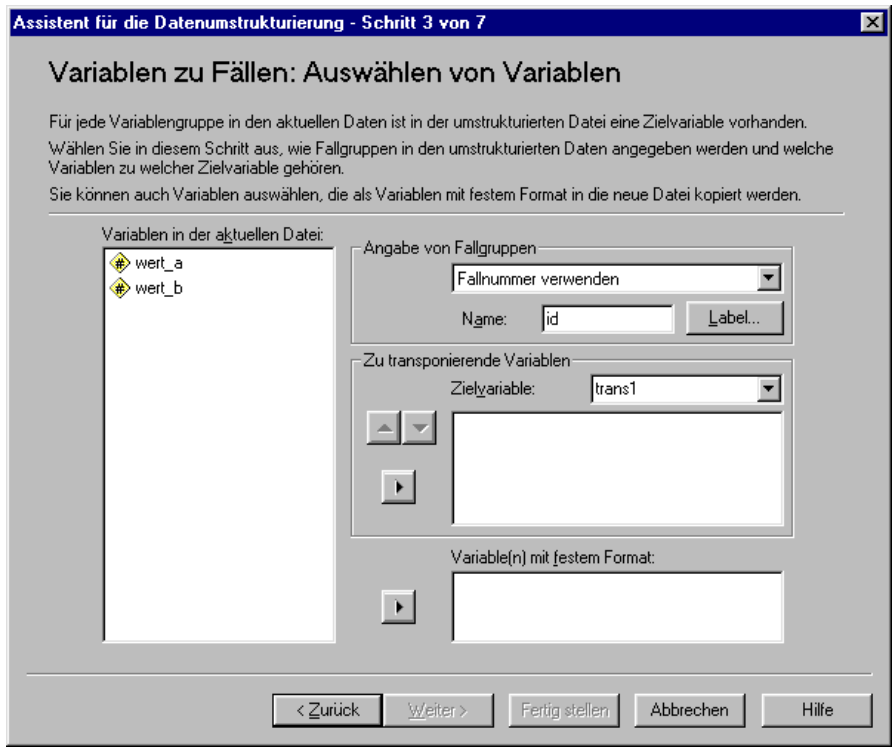
- **Eine.** Der Assistent erstellt in der neuen Datei eine einzige umstrukturierte Variable aus einer Variablengruppe der aktuellen Datei.
- **Mehrere.** Der Assistent erstellt in der neuen Datei mehrere umstrukturierte Variablen. Die angegebene Anzahl wirkt sich auf den nächsten Schritt aus, in dem der Assistent automatisch die angegebene Anzahl von neuen Variablen erstellt.

Assistent für die Datenumstrukturierung (Variablen zu Fällen): Auswählen Variablen

Anmerkung: Der Assistent zeigt diesen Schritt an, wenn Variablengruppen in Zeilen umstrukturiert werden sollen.

In diesem Schritt geben Sie an, wie die Variablen der aktuellen Datei in der neuen Datei verwendet werden sollen. Sie können auch eine Variable erstellen, welche die Zeilen in der neuen Datei identifiziert.

Abbildung 9-22
Assistent für die Datenumstrukturierung: Variablen auswählen, Schritt 3


Assistent für die Datenumstrukturierung - Schritt 3 von 7
Variablen zu Fällen: Auswählen von Variablen
Für jede Variablengruppe in den aktuellen Daten ist in der umstrukturierten Datei eine Zielvariable vorhanden.
Wählen Sie in diesem Schritt aus, wie Fallgruppen in den umstrukturierten Daten angegeben werden und welche Variablen zu welcher Zielvariable gehören.
Sie können auch Variablen auswählen, die als Variablen mit festem Format in die neue Datei kopiert werden.
Variablen in der aktuellen Datei:
wert_a
wert_b
Angabe von Fallgruppen
Fallnummer verwenden
Name:
id
Label...
Zu transponierende Variablen
Zielvariable:
trans1
Variable(n) mit festem Format:
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

Wie sollen die neuen Zeilen angegeben werden? Sie können in der neuen Datei eine Variable erstellen, welche die Zeile in der aktuellen Datei identifiziert, die zum Erstellen einer Gruppe von neuen Zeilen verwendet wurde. Diese Bezeichnervariable kann aus fortlaufenden Fallnummern oder den Werten der Variablen bestehen. Verwenden Sie die Steuerelemente in “Angabe von Fallgruppen”, um die Bezeichnervariable in der neuen Datei zu definieren. Klicken Sie auf eine Zelle, um den vorgegebenen Variablennamen zu ändern und ein beschreibendes Variablenlabel für die Bezeichnervariable einzugeben.

Was soll in der neuen Datei umstrukturiert werden? Im vorhergehenden Schritt haben Sie angegeben, wie viele Variablengruppen umstrukturiert werden sollen. Der Assistent erstellt für jede Gruppe eine neue Variable. Die Werte für die Variablengruppe werden in dieser Variablen in der neuen Datei dargestellt. Verwenden Sie die Steuerelemente in "Zu transponierende Variablen", um die umstrukturierte Variable in der neuen Datei zu definieren.

So geben Sie eine umstrukturierte Variable an:

- Fügen Sie die Variablen der zu transformierenden Variablengruppe der Liste "Zu transponierende Variablen" hinzu. Alle Variablen der Gruppe müssen vom selben Typ sein (numerische oder String-Variablen).

  Dieselbe Variable kann mehrfach in der Variablengruppe vorhanden sein (Variablen werden in der Regel kopiert und nicht aus der Liste der Quellvariablen verschoben). Die Werte werden in der neuen Datei wiederholt.

So geben Sie mehrere umstrukturierte Variablen an:

- Wählen Sie in der Dropdown-Liste "Zielvariable" die erste Zielvariable aus, die Sie definieren möchten.
- Fügen Sie die Variablen der zu transformierenden Variablengruppe der Liste "Zu transponierende Variablen" hinzu. Alle Variablen der Gruppe müssen vom selben Typ sein (numerische oder String-Variablen). Eine Variable kann mehr als einmal in der Variablengruppe enthalten sein. (Variablen werden aus der Liste der Quellvariablen nicht verschoben, sondern kopiert, und ihre Werte werden in der neuen Datei wiederholt.)
- Wählen Sie die nächste Zielvariable aus, die Sie definieren möchten, und wiederholen Sie den Auswahlvorgang für alle verfügbaren Zielvariablen.
  - Obwohl eine Variable mehrfach in einer Zielvariablengruppe vorkommen kann, darf dieselbe Variable nicht in mehreren Zielvariablengruppen beinhaltet sein.
  - Jede Liste von Zielvariablengruppen muß die gleiche Anzahl von Variablen enthalten. (Mehrmals aufgeführte Variablen werden in die Zählung einbezogen.)
  - Die Anzahl der Zielvariablengruppen wird von der Anzahl der im vorhergehenden Schritt angegebenen Variablengruppen bestimmt. Die Standardvariablennamen können hier geändert werden, aber um die Anzahl der zu restrukturierenden

Variablengruppen zu ändern, müssen Sie zum vorhergehenden Schritt zurückkehren.

- Bevor Sie mit dem nächsten Schritt fortfahren können, müssen Sie für alle Zielvariablen eine Variablengruppe definiert haben (durch Auswählen von Variablen in der Quelliste).

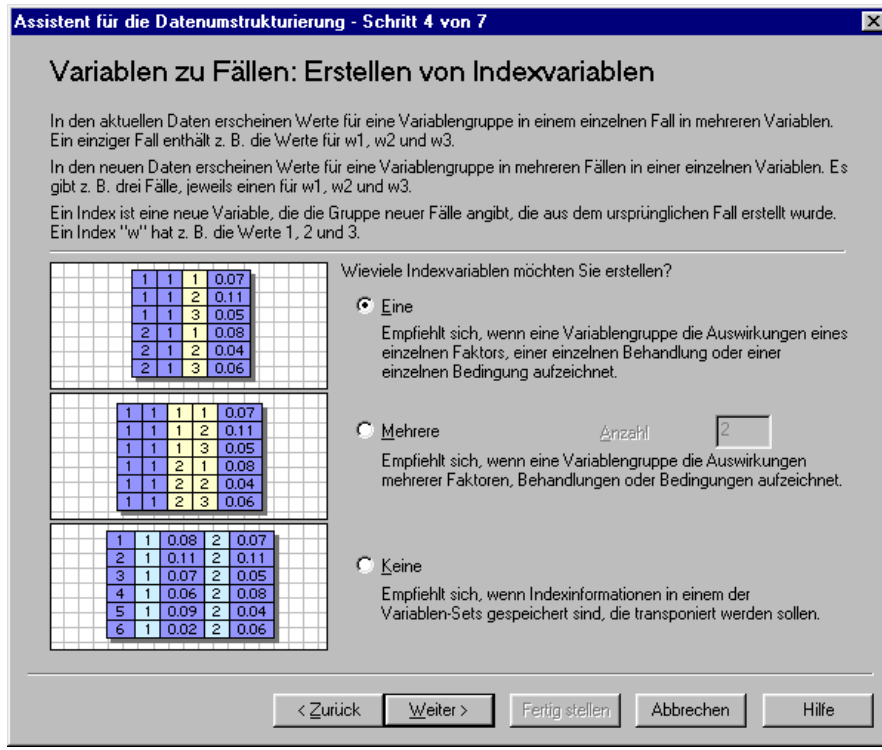
Was soll in die neue Datei kopiert werden? Variablen, die nicht umstrukturiert werden, können in die neue Datei kopiert werden. Die Werte für diese Variablen werden in die neuen Zeilen übertragen. Verschieben Sie die Variablen, die in die neue Datei kopiert werden sollen, in die Liste "Variable(n) mit festem Format".

Assistent für die Datenumstrukturierung (Variablen zu Fällen): Erstellen von Indexvariablen

Anmerkung: Der Assistent zeigt diesen Schritt an, wenn Variablengruppen in Zeilen umstrukturiert werden sollen.

In diesem Schritt legen Sie fest, ob Indexvariablen erstellt werden sollen. Bei einem Index handelt es sich um eine neue Variable, die eine Zeilengruppe fortlaufend anhand der ursprünglichen Variablen identifiziert, aus der die neue Zeile erstellt wurde.

Abbildung 9-23
Assistent für die Datenumstrukturierung: Erstellen von Indexvariablen, Schritt 4


Assistent für die Datenumstrukturierung - Schritt 4 von 7
Variablen zu Fällen: Erstellen von Indexvariablen
In den aktuellen Daten erscheinen Werte für eine Variablengruppe in einem einzelnen Fall in mehreren Variablen. Ein einziger Fall enthält z. B. die Werte für w1, w2 und w3.
In den neuen Daten erscheinen Werte für eine Variablengruppe in mehreren Fällen in einer einzelnen Variablen. Es gibt z. B. drei Fälle, jeweils einen für w1, w2 und w3.
Ein Index ist eine neue Variable, die die Gruppe neuer Fälle angibt, die aus dem ursprünglichen Fall erstellt wurde. Ein Index "w" hat z. B. die Werte 1, 2 und 3.
Wieviele Indexvariablen möchten Sie erstellen?
Eine
Empfiehlt sich, wenn eine Variablengruppe die Auswirkungen eines einzelnen Faktors, einer einzelnen Behandlung oder einer einzelnen Bedingung aufzeichnet.
Mehrere
Anzahl
2
Empfiehlt sich, wenn eine Variablengruppe die Auswirkungen mehrerer Faktoren, Behandlungen oder Bedingungen aufzeichnet.
Keine
Empfiehlt sich, wenn Indexinformationen in einem der Variablen-Sets gespeichert sind, die transponiert werden sollen.
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

Wie viele Indexvariablen soll die neue Datei enthalten? Indexvariablen können in Prozeduren von SPSS als Gruppenvariablen verwendet werden. In den meisten Fällen ist eine einzige Indexvariable ausreichend. Wenn die Variablengruppen in der aktuellen Datei allerdings mehrere Faktorstufen darstellen, sind unter Umständen mehrere Indizes erforderlich.

- **Eine.** Der Assistent erstellt eine einzige Indexvariable.
- **Mehrere.** Der Assistent erstellt mehrere Indizes. Geben Sie die Anzahl der zu erstellenden Indizes ein. Die angegebene Anzahl wirkt sich auf den nächsten Schritt aus, in dem der Assistent automatisch die angegebene Anzahl von Indizes erstellt.
- **Keine.** Wählen Sie diese Option aus, wenn in der neuen Datei keine Indexvariable erstellt werden soll.

Beispiel für einen Index bei der Umstrukturierung von Variablen zu Fällen

Die aktuellen Daten enthalten eine Variablengruppe für die *Breite* und einen Faktor für die *Zeit*. Die Breite wurde dreimal gemessen und in *w1*, *w2* und *w3* aufgezeichnet.

Abbildung 9-24
Aktuelle Daten für einen Index

	subjekt	w1	w2	w3
1	1	6.70	4.30	5.70
2	2	7.10	5.90	5.60

Die Variablengruppe wird nun in eine einzelne Variable für *Breite* umstrukturiert. Zudem wird ein einzelner numerischer Index erstellt. In der folgenden Tabelle werden die neuen Daten abgebildet.

Abbildung 9-25
Neue umstrukturierte Daten mit einem Index

	subjekt	index	breite
1	1	1	6.70
2	1	2	4.30
3	1	3	5.70
4	2	1	7.10
5	2	2	5.90
6	2	3	5.60

Der *Index* beginnt bei 1 und zählt jede Variable in der Gruppe in aufsteigender Reihenfolge. Er beginnt jedesmal neu, wenn in der Originaldatei eine neue Zeile gefunden wird. Der *Index* kann von nun an in SPSS-Prozeduren verwendet werden, für die eine Gruppenvariable erforderlich ist.

Beispiel für zwei Indizes bei der Umstrukturierung von Variablen zu Fällen

Wenn eine Variablengruppe mehrere Faktoren aufzeichnet, können Sie mehrere Indizes erstellen. Die aktuellen Daten müssen dabei jedoch so angeordnet sein, daß die Stufen des ersten Faktors einen Primärindex bilden, der von allen weiteren Faktorstufen durchlaufen wird. Die aktuelle Datei enthält eine Variablengruppe für die *Breite* und zwei Faktoren, *A* und *B*. Die Daten sind so angeordnet, daß die Stufen von Faktor *B* die Stufen von Faktor *A* durchlaufen.

Abbildung 9-26
Aktuelle Daten für zwei Indizes

	subjekt	w_a1b1	w_a1b2	w_a2b1	w_a2b2
1	1	5.50	6.40	5.80	5.90
2	2	7.40	7.10	5.60	6.70

Die Variablengruppe wird nun in eine einzelne Variable für die *Breite* umstrukturiert, und dabei werden zwei Indizes erstellt. In der folgenden Tabelle werden die neuen Daten abgebildet.

Abbildung 9-27
Neue, umstrukturierte Daten mit zwei Indizes

	subjekt	index_a	index_b	breite
1	1	1	1	5.50
2	1	1	2	6.40
3	1	2	1	5.80
4	1	2	2	5.90
5	2	1	1	7.40
6	2	1	2	7.10
7	2	2	1	5.60
8	2	2	2	6.70

Assistent für die Datenumstrukturierung (Variablen zu Fällen): Erstellen einer Indexvariablen

Anmerkung: Der Assistent zeigt diesen Schritt an, wenn Sie die Umstrukturierung von Variablengruppen in Zeilen und die Erstellung von einer Indexvariablen auswählen.

In diesem Schritt legen Sie die gewünschten Werte für die Indexvariable fest. Bei den Werten kann es sich um fortlaufende Nummern oder um die Namen der Variablen aus der ursprünglichen Variablengruppe handeln. Außerdem können Sie einen Namen und ein Label für die neue Indexvariable angeben.

Abbildung 9-28
Assistent für die Datenumstrukturierung: Erstellen einer Indexvariablen, Schritt 5

Für weitere Informationen siehe “Beispiel für einen Index bei der Umstrukturierung von Variablen zu Fällen” auf S. 257.

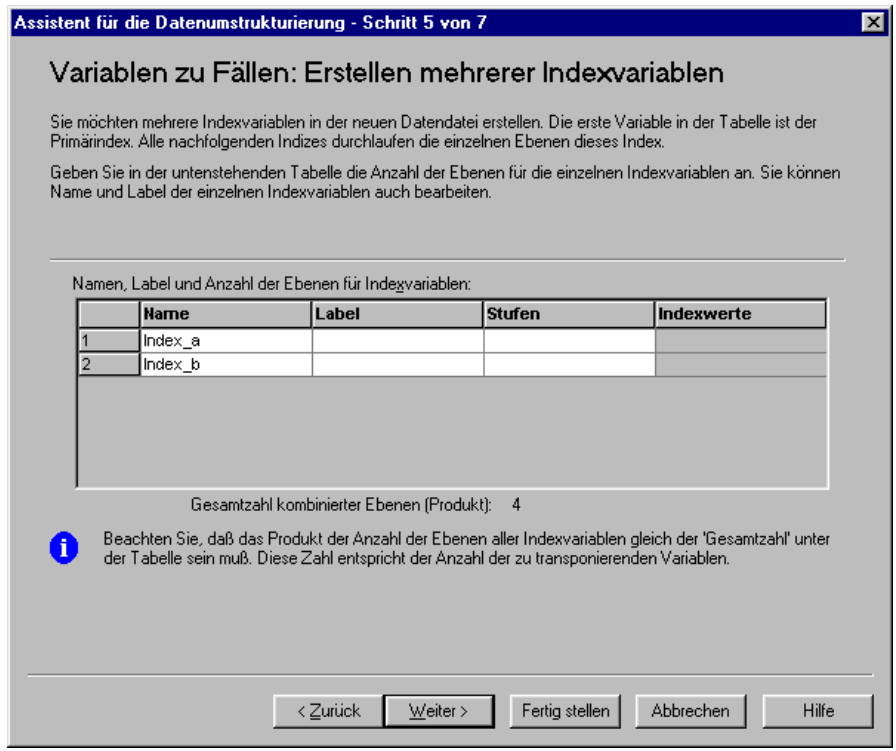
- **Fortlaufende Zahlen.** Der Assistent weist automatisch fortlaufende Nummern als Indexwerte zu.
- **Variablennamen.** Der Assistent verwendet die Namen der ausgewählten Variablengruppe als Indexwerte. Wählen Sie eine Variablengruppe in der Liste aus.
- **Namen und Labels.** Klicken Sie auf eine Zelle, um den vorgegebenen Variablennamen zu ändern und ein beschreibendes Variablenlabel für die Indexvariable einzugeben.

Assistent für die Datenumstrukturierung (Variablen zu Fällen): Erstellen mehrerer Indexvariablen

Anmerkung: Der Assistent zeigt diesen Schritt an, wenn Variablengruppen in Zeilen umstrukturiert und mehrere Indexvariablen erstellt werden sollen.

In diesem Schritt geben Sie die Anzahl der Stufen für jede Indexvariable an. Außerdem können Sie einen Namen und ein Label für die neue Indexvariable angeben.

Abbildung 9-29
Assistent für die Datenumstrukturierung: Erstellen mehrerer Indexvariablen, Schritt 5


Assistent für die Datenumstrukturierung - Schritt 5 von 7
Variablen zu Fällen: Erstellen mehrerer Indexvariablen
Sie möchten mehrere Indexvariablen in der neuen Datendatei erstellen. Die erste Variable in der Tabelle ist der Primärindex. Alle nachfolgenden Indizes durchlaufen die einzelnen Ebenen dieses Index.
Geben Sie in der untenstehenden Tabelle die Anzahl der Ebenen für die einzelnen Indexvariablen an. Sie können Name und Label der einzelnen Indexvariablen auch bearbeiten.
Namen, Label und Anzahl der Ebenen für Indexvariablen:
Name
Label
Stufen
Indexwerte
1
Index_a
2
Index_b
Gesamtzahl kombinierter Ebenen (Produkt): 4
Beachten Sie, daß das Produkt der Anzahl der Ebenen aller Indexvariablen gleich der 'Gesamtzahl' unter der Tabelle sein muß. Diese Zahl entspricht der Anzahl der zu transponierenden Variablen.
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

Für weitere Informationen siehe “Beispiel für zwei Indizes bei der Umstrukturierung von Variablen zu Fällen” auf S. 257.

Wie viele Stufen sind in der aktuellen Datei aufgezeichnet? Überprüfen Sie, wie viele Faktorstufen in den aktuellen Daten aufgezeichnet sind. Eine **Stufe** definiert eine Gruppen von Fällen, die identischen Bedingungen ausgesetzt wurden. Wenn mehrere

Faktoren vorhanden sind, müssen die aktuellen Daten so angeordnet sein, daß die Stufen des ersten Faktors einen Primärindex darstellen, der von allen weiteren Faktorstufen durchlaufen wird.

Wie viele Stufen soll die neue Datei enthalten? Geben Sie für jeden Index die Anzahl der Stufen ein. Die Werte für mehrere Indexvariablen sind immer fortlaufende Nummern. Die Werte beginnen bei 1 und werden für jede Stufe erhöht. Die Schrittgröße für die Erhöhung der Indizes ist für den ersten Index am kleinsten und für den letzten Index am größten.

Gesamtzahl kombinierter Ebenen. Es können nur so viele Stufen erstellt werden, wie in den aktuellen Daten vorhanden sind. Da die umstrukturierten Daten eine Zeile pro Behandlungskombination enthalten, überprüft der Assistent die Anzahl der zu erstellenden Stufen. Er vergleicht das Produkt der von Ihnen erstellten Stufen mit der Anzahl der Variablen in der Variablengruppe. Die beiden Werte müssen übereinstimmen.

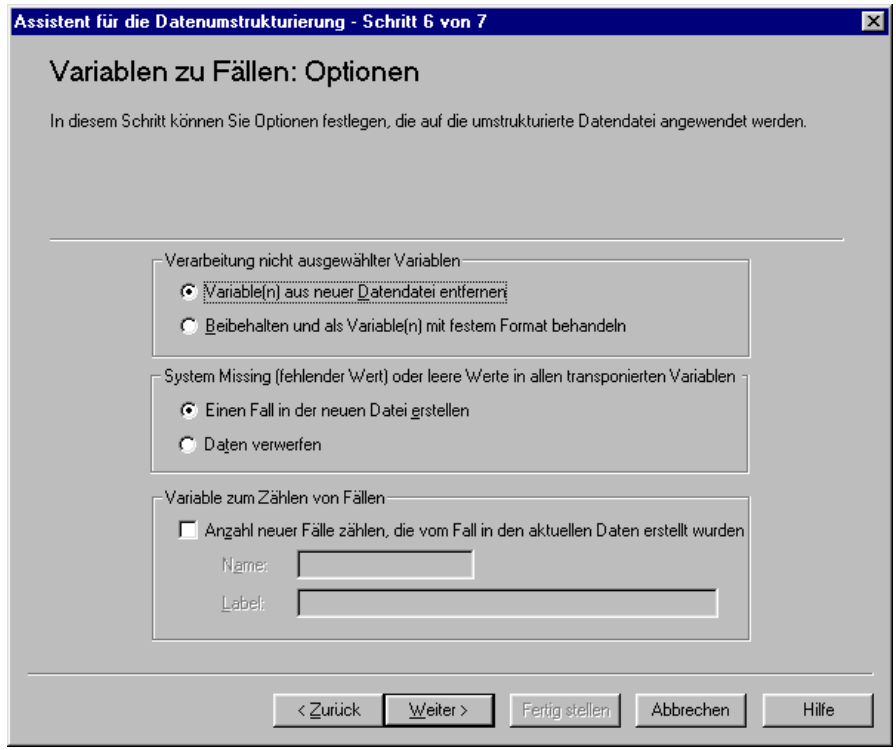
Namen und Labels. Klicken Sie auf eine Zelle, um den vorgegebenen Variablennamen zu ändern und ein beschreibendes Variablenlabel für die Indexvariablen einzugeben.

Assistent für die Datenumstrukturierung (Variablen zu Fällen): Optionen

Anmerkung: Der Assistent zeigt diesen Schritt an, wenn Variablengruppen in Zeilen umstrukturiert werden sollen.

In diesem Schritt legen Sie die Optionen für die neue umstrukturierte Datei fest.

Abbildung 9-30
Assistent für die Datenumstrukturierung: Optionen, Schritt 6


Assistent für die Datenumstrukturierung - Schritt 6 von 7
Variablen zu Fällen: Optionen
In diesem Schritt können Sie Optionen festlegen, die auf die umstrukturierte Datendatei angewendet werden.
Verarbeitung nicht ausgewählter Variablen
Variable(n) aus neuer Datendatei entfernen
Beibehalten und als Variable(n) mit festem Format behandeln
System Missing (fehlender Wert) oder leere Werte in allen transponierten Variablen
Einen Fall in der neuen Datei erstellen
Daten verwerfen
Variable zum Zählen von Fällen
Anzahl neuer Fälle zählen, die vom Fall in den aktuellen Daten erstellt wurden
Name:
Label:
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

Sollen nicht ausgewählte Variablen verworfen werden? Im Schritt "Variablen auswählen" (Schritt 3) haben Sie aus den aktuellen Daten die umzustrukturierenden Variablengruppen, die zu kopierenden Variablen und eine Bezeichnervariable ausgewählt. Die Daten der ausgewählten Variablen werden in die neue Datei übertragen. Wenn die aktuellen Daten weitere Variablen enthalten, können Sie festlegen, ob diese verworfen oder beibehalten werden sollen.

Sollen fehlende Daten beibehalten werden? Der Assistent überprüft jede potentiell neue Zeile auf das Vorhandensein von Null-Werten. Ein **Null-Wert** ist ein systemdefinierter fehlender oder leerer Wert. Sie können festlegen, ob Zeilen mit Null-Werten beibehalten oder verworfen werden sollen.

Soll eine Zählvariable erstellt werden? Der Assistent kann in der neuen Datei eine **Zählvariable** erstellen. Diese enthält die Anzahl der neuen Zeilen, die von einer Zeile in den aktuellen Daten erzeugt wurden. Eine Zählvariable kann nützlich sein, wenn Sie die Null-Werte in der neuen Datei verwerfen möchten, da in diesem Fall für eine gegebene Zeile in den aktuellen Daten die Anzahl der neu zu erstellenden Zeilen variieren kann. Klicken Sie auf eine Zelle, um den vorgegebenen Variablennamen zu ändern und ein beschreibendes Variablenlabel für die Zählvariable einzugeben.

Assistent für die Datenumstrukturierung (Fälle zu Variablen): Auswählen Variablen

Anmerkung: Der Assistent zeigt diesen Schritt an, wenn Fallgruppen in Spalten umstrukturiert werden sollen.

In diesem Schritt geben Sie an, wie die Variablen der aktuellen Datei in der neuen Datei verwendet werden sollen.

Abbildung 9-31
Assistent für die Datenumstrukturierung: Variablen auswählen, Schritt 2

Assistent für die Datenumstrukturierung - Schritt 2 von 5
Fälle zu Variablen: Auswählen von Variablen
Daten aus Fallgruppen in der aktuellen Datei werden in der neuen Datei in Einzelfälle umstrukturiert.
Wählen Sie Variablen aus, die Fallgruppen angeben, indem Sie sie in die Liste "Bezeichnervariable(n)" verschieben. Sie können auch "Indexvariable(n)" auswählen.
Die Variablen, die in der aktuellen Datei in der Liste der Variablen verbleiben, enthalten entweder Daten, die in einer Fallgruppe variieren, oder Daten, die nicht variieren.
Eine Variable mit variierenden Daten wird in der umstrukturierten Datei zu einer Gruppe von neuen Variablen. Eine Variable mit nicht variierenden Daten wird in die neue Datei kopiert.
Variablen in der aktuellen Datei:
wert
Bezeichnervariable(n):
id
Indexvariable(n):
zeit
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

Wodurch werden Fallgruppen in den aktuellen Daten bezeichnet? Bei einer **Fallgruppe** handelt es sich um eine Gruppe von Zeilen, die aufeinander bezogen sind, weil sie dieselbe Beobachtungseinheit messen, beispielsweise eine Person oder eine Institution. Sie müssen im Assistenten angeben, welche Variablen der aktuellen Datei die Fallgruppen identifizieren, damit die einzelnen Fallgruppen in der neuen Datei jeweils in eine einzelne Zeile konsolidiert werden können. Verschieben Sie die Variablen, die in der aktuellen Datei Fallgruppen identifizieren, in die Liste “Bezeichnervariable(n)”. Variablen, die zum Aufteilen der aktuellen Datendatei dienen, werden automatisch zum Identifizieren von Fallgruppen verwendet. Bei jedem Auftreten einer neuen Kombination von Identifizierungswerten wird eine neue Zeile erstellt, so daß die Fälle in der aktuellen Datei nach den Werten der Bezeichnervariablen sortiert sein müssen, und zwar in der Reihenfolge, in der die

Variablen in der Liste “Bezeichnervariable(n)” aufgeführt sind. Wenn die aktuelle Datendatei noch nicht sortiert ist, können Sie dies im nächsten Schritt nachholen.

Auf welche Weise sollen die neuen Variablengruppen in der neuen Datei erstellt werden? In den ursprünglichen Daten wird eine Variable in einer einzelnen Spalte aufgeführt. In den neuen Datendatei wird diese Variable in mehreren Spalten aufgelistet. **Indexvariablen** sind Variablen in den aktuellen Daten, die vom Assistenten zum Erstellen der neuen Spalten verwendet werden. Die umstrukturierten Daten enthalten eine neue Variable für jeden eindeutigen Wert in diesen Spalten. Verschieben Sie die Variablen, die zur Erstellung der neuen Variablengruppen verwendet werden sollen, in die Liste “Indexvariable(n)”. Sie können die neuen Spalten auch nach Indizes ordnen, wenn die entsprechenden Optionen vom Assistenten anzeigt werden.

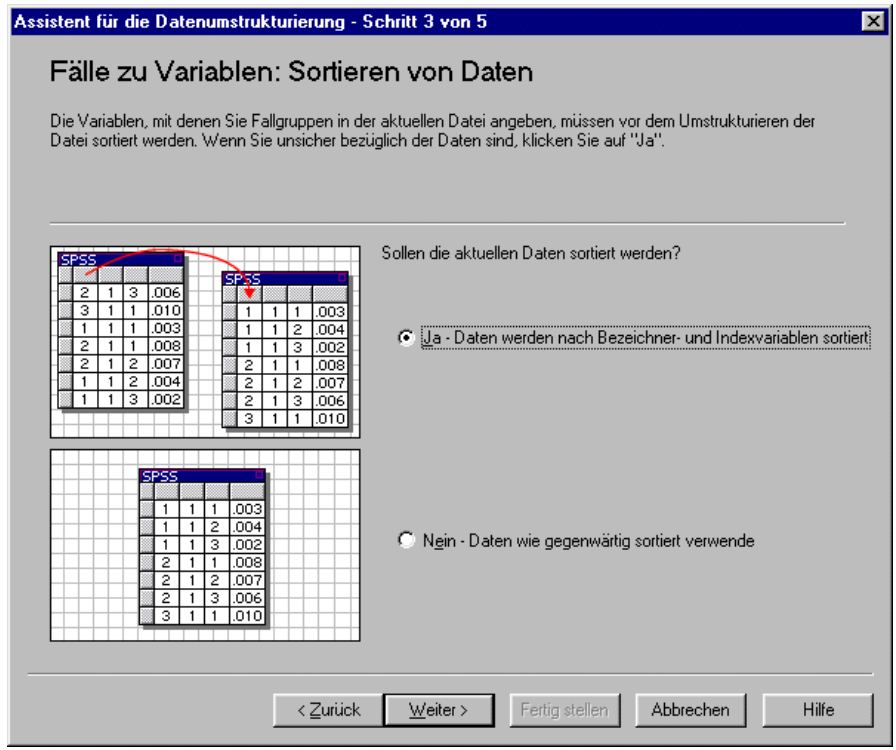
Was passiert mit den anderen Spalten? Der Assistent entscheidet automatisch, was mit den Variablen geschieht, die in der Liste “Aktuelle Datei” verbleiben. Er überprüft jede Variable, um festzustellen, ob die Datenwerte innerhalb einer Fallgruppe variieren. Wenn dies der Fall ist, strukturiert der Assistent die Werte in eine Variablengruppe in der neuen Datei um. Wenn dies nicht der Fall ist, kopiert der Assistent die Werte in die neue Datei.

Assistent für die Datenumstrukturierung (Fälle zu Variablen): Sortierung. Daten

Anmerkung: Der Assistent zeigt diesen Schritt an, wenn Fallgruppen in Spalten umstrukturiert werden sollen.

In diesem Schritt legen Sie fest, ob die Daten in der aktuellen Datei vor der Umstrukturierung sortiert werden sollen. Bei jedem Auftreten einer neuen Kombination von Identifizierungswerten wird vom Assistenten eine neue Zeile erstellt. Aus diesem Grund ist es wichtig, daß die Daten nach den Variablen sortiert sind, die die Fallgruppen identifizieren.

Abbildung 9-32
Assistent für die Datenumstrukturierung: Sortieren von Daten, Schritt 3


Assistent für die Datenumstrukturierung - Schritt 3 von 5
Fälle zu Variablen: Sortieren von Daten
Die Variablen, mit denen Sie Fallgruppen in der aktuellen Datei angeben, müssen vor dem Umstrukturieren der Datei sortiert werden. Wenn Sie unsicher bezüglich der Daten sind, klicken Sie auf "Ja".
Sollen die aktuellen Daten sortiert werden?
Ja - Daten werden nach Bezeichner- und Indexvariablen sortiert
Nein - Daten wie gegenwärtig sortiert verwende
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

Wie sind die Zeilen in der aktuellen Datei geordnet? Überprüfen Sie, wie die aktuellen Daten sortiert sind und welche (im vorhergehenden Schritt angegebenen) Variablen zum Identifizieren von Fallgruppen verwendet werden.

- **Ja.** Der Assistent sortiert die aktuellen Daten automatisch nach den Bezeichnervariablen, und zwar in der Reihenfolge, in der die Variablen im vorhergehenden Schritt in die Liste “Bezeichnervariable(n)” eingetragen wurden. Wählen Sie diese Option aus, wenn die Daten noch nicht nach Bezeichnervariablen sortiert sind oder Sie sich nicht sicher sind. Diese Option

erfordert einen zusätzlichen Datendurchlauf, aber sie garantiert, daß die Zeilen für die Umstrukturierung korrekt sortiert sind.

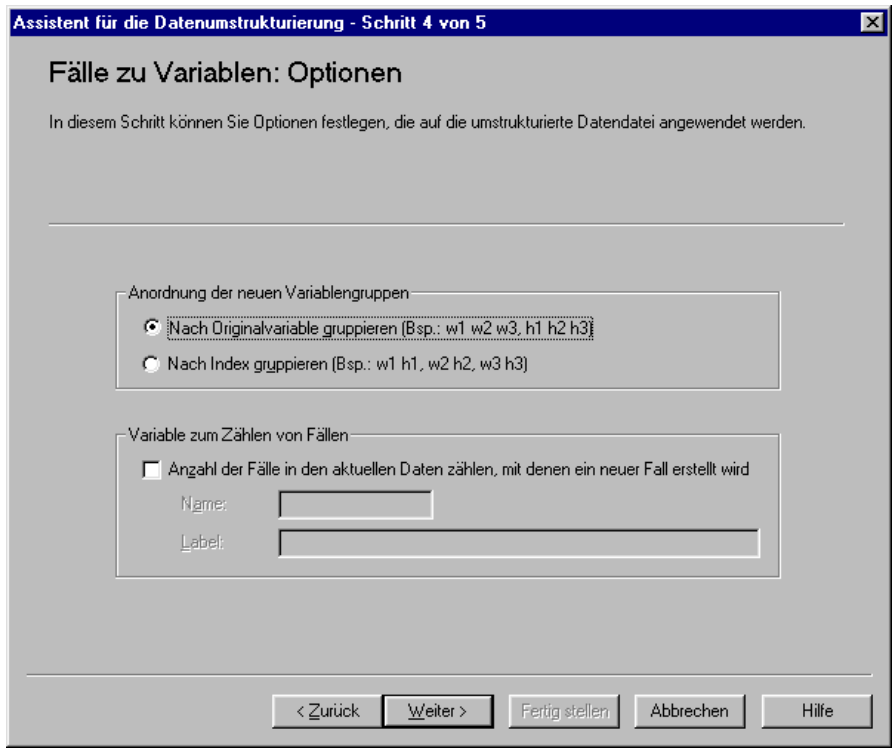
- **Nein.** Der Assistent sortiert die aktuellen Daten nicht. Wählen Sie diese Option aus, wenn Sie sicher sind, daß die aktuellen Daten bereits nach den Variablen sortiert sind, die die Fallgruppen identifizieren.

Assistent für die Datenumstrukturierung (Fälle zu Variablen): Optionen

Anmerkung: Der Assistent zeigt diesen Schritt an, wenn Fallgruppen in Spalten umstrukturiert werden sollen.

In diesem Schritt legen Sie die Optionen für die neue umstrukturierte Datei fest.

Abbildung 9-33
Assistent für die Datenumstrukturierung: Optionen, Schritt 4


Assistent für die Datenumstrukturierung - Schritt 4 von 5
Fälle zu Variablen: Optionen
In diesem Schritt können Sie Optionen festlegen, die auf die umstrukturierte Datendatei angewendet werden.
Anordnung der neuen Variablengruppen
Nach Originalvariable gruppieren (Bsp.: w1 w2 w3, h1 h2 h3)
Nach Index gruppieren (Bsp.: w1 h1, w2 h2, w3 h3)
Variable zum Zählen von Fällen
Anzahl der Fälle in den aktuellen Daten zählen, mit denen ein neuer Fall erstellt wird
Name:
Label:
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

Wie sollen die neuen Variablengruppen in der neuen Datei sortiert werden?

- **Nach Variablen.** Der Assistent gruppiert die neuen Variablen nach den ursprünglichen Variablen, aus denen sie erstellt wurden.
- **Nach Index.** Der Assistent gruppiert die Variablen nach den Werten der Indexvariablen.

Beispiel. Die umzustrukturierenden Variablen sind *w* und *h*, und der Index lautet *Monat*:

w	h	Monat

Die Gruppierung nach Variablen ergibt das folgende Ergebnis:

w.jan	w.feb	h.jan

Die Gruppierung nach Indizes ergibt das folgende Ergebnis:

w.jan	h.jan	w.feb

Soll eine Zählvariable erstellt werden? Der Assistent kann in der neuen Datei eine Zählvariable erstellen. Diese Zählvariable enthält die Anzahl der Zeilen in den aktuellen Daten, die zur Erstellung einer Zeile in der neuen Datendatei verwendet wurden.

Sollen Indikatorvariablen erstellt werden? Mit Hilfe der Indexvariablen kann der Assistent in der neuen Datendatei **Indikatorvariablen** erstellen. Er erstellt für jeden eindeutigen Wert der Indexvariable eine neue Variable. Die Indikatorvariablen geben das Vorhandensein bzw. Nichtvorhandensein eines Wertes für einen Fall an. Wenn der Fall einen Wert aufweist, besitzt eine Indikatorvariable den Wert 1, andernfalls besitzt sie den Wert 0.

Beispiel. Die Indexvariable ist *Produkt*. Sie dient zum Aufzeichnen der von einem Kunden erworbenen Produkte. Die ursprünglichen Daten lauten wie folgt:

Kunde	**Produkt**
1	Huhn
1	Eier

Kunde	Produkt
2	Eier
3	Huhn

Beim Erstellen einer Indikatorvariable entsteht eine neue Variable für jeden eindeutigen Wert von *Produkt*. Die umstrukturierten Daten lauten wie folgt:

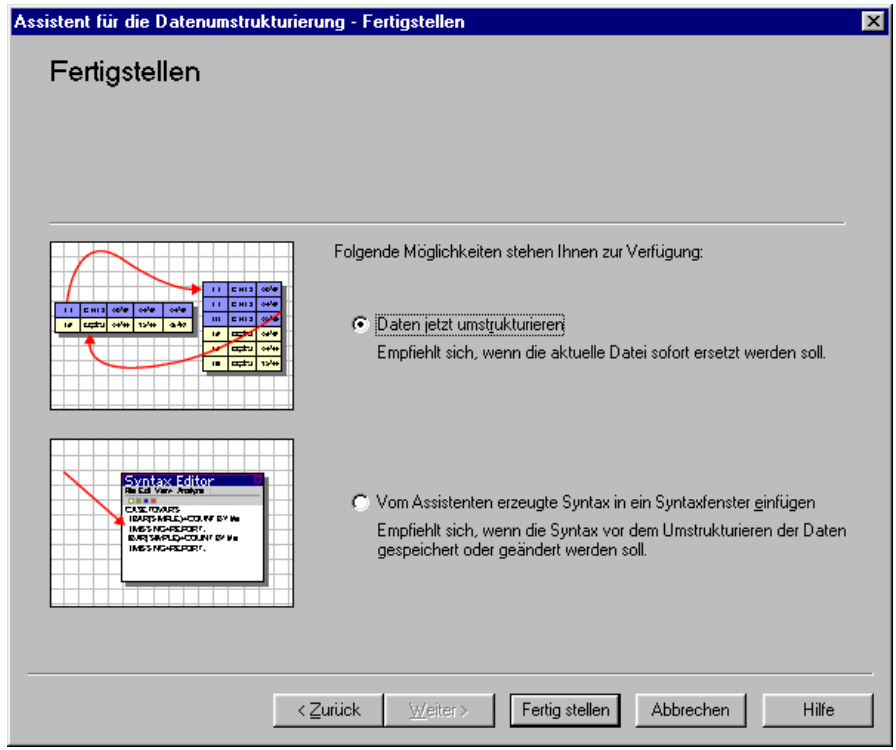
Kunde	indHuhn	indEier
1	1	1
2	0	1
3	1	0

In diesem Beispiel können die umstrukturierten Daten zum Ermitteln der Häufigkeiten für die von Kunden erworbenen Produkte verwendet werden.

Assistent für die Datenumstrukturierung: Fertigstellen

Dies ist der letzte Schritt im Assistenten für die Datenumstrukturierung. Geben Sie an, welcher Vorgang mit den von Ihnen getroffenen Angaben ausgeführt werden soll.

Abbildung 9-34
Assistent für die Datenumstrukturierung: Fertigstellen


Assistent für die Datenumstrukturierung - Fertigstellen
Fertigstellen
Folgende Möglichkeiten stehen Ihnen zur Verfügung:
Daten jetzt umstrukturieren
Empfiehlt sich, wenn die aktuelle Datei sofort ersetzt werden soll.
Syntax Editor
Vom Assistenten erzeugte Syntax in ein Syntaxfenster einfügen
Empfiehlt sich, wenn die Syntax vor dem Umstrukturieren der Daten gespeichert oder geändert werden soll.
< Zurück
Weiter >
Fertig stellen
Abbrechen
Hilfe

- **Daten jetzt umstrukturieren.**Der Assistent erstellt die neue umstrukturierte Datei. Wählen Sie diese Option aus, wenn die aktuelle Datei sofort ersetzt werden soll. *Anmerkung*: Wenn die ursprünglichen Daten gewichtet sind, sind auch die neuen Daten gewichtet, es sei denn, die für die Gewichtung verwendete Variable wird in der neuen Datei umstrukturiert oder verworfen.
- **Vom Assistenten erzeugte Syntax in ein Syntaxfenster einfügen**. Der Assistent fügt die von ihm erstellte Syntax in ein Syntax-Fenster ein. Wählen Sie diese Option aus, wenn die aktuelle Datei noch nicht ersetzt werden soll, oder wenn die Syntax noch weitergehend bearbeitet bzw. zur späteren Verwendung gespeichert werden soll.

Kapitel
10

Arbeiten mit Ausgaben

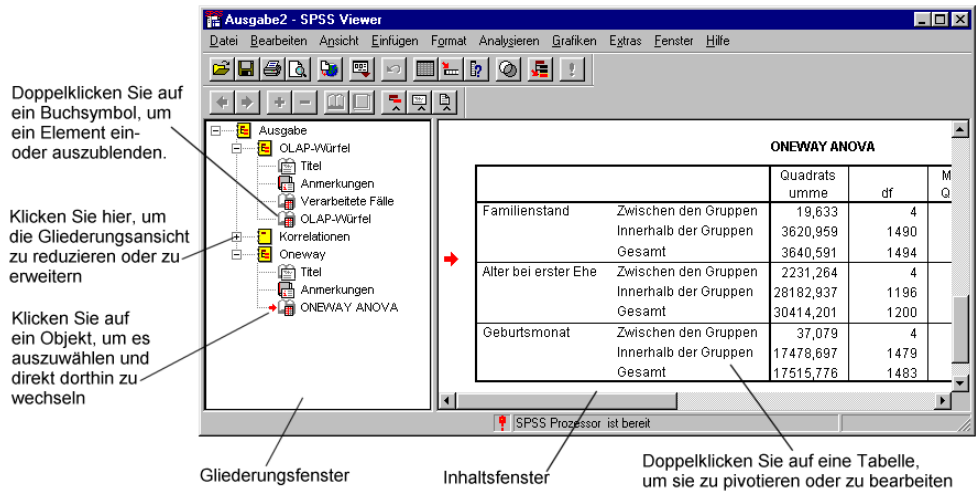
Nach dem Ausführen einer Prozedur werden die Ergebnisse in einem Fenster angezeigt, das “Viewer” heißt. In diesem Fenster können Sie problemlos zwischen den verschiedenen Teilen der Ausgabe wechseln. Außerdem können Sie die Ausgaben bearbeiten und so Dokumente erstellen, die genau die gewünschten Ausgaben enthalten.

Viewer

Die Ergebnisse werden im Viewer angezeigt. Sie können den Viewer für folgende Vorgänge verwenden:

- Durchsuchen der Ergebnisse,
- Ein- und Ausblenden von ausgewählten Tabellen und Diagrammen,
- Ändern der Anzeigereihenfolge der Ergebnisse durch Verschieben ausgewählter Objekte,
- Verschieben von Objekten zwischen dem Viewer und anderen Anwendungen.

Abbildung 10-1
Viewer



Der Viewer ist in zwei Fensterbereiche aufgeteilt:

- Der linke Fensterbereich des Viewers enthält eine Gliederungsansicht des Inhalts.
- Der rechte Fensterbereich enthält Statistiktabellen, Diagramme und Textausgabe.

Sie können die Bildlaufleisten verwenden, um die Ergebnisse zu sichten. Sie können aber auch auf Objekte in der Gliederung klicken, um direkt zu den entsprechenden Tabellen oder Diagrammen zu wechseln. Wenn Sie die Breite des Gliederungfensters ändern möchten, können Sie auf dessen rechten Rahmen klicken und ihn mit gedrückter Maustaste auf die gewünschte Breite ziehen.

Arbeiten mit dem Text-Viewer

Falls Sie lieber mit einfachen Textausgaben als mit interaktiven Pivot-Tabellen arbeiten, verwenden Sie hierfür den Text-Viewer.

So rufen Sie den Text-Viewer auf:

▶ Wählen Sie in einem beliebigen Fenster die folgenden Befehle aus den Menüs aus:

Bearbeiten
  Optionen...

▶ Klicken Sie auf der Registerkarte “Allgemein” unter Ausgabetyp auf Text.

- Sie können die Formatoptionen für die Ausgabe des Text-Viewers ändern, indem Sie auf die Registerkarte Text-Viewer klicken und dort die entsprechenden Einstellungen vornehmen.

  Für weitere Informationen siehe “Text-Viewer” in Kapitel 11 auf S. 307. Außerdem können Sie die Hilfe nach weiteren Informationen durchsuchen:

- Wählen Sie in einem beliebigen Fenster die folgenden Befehle aus den Menüs aus:

  Hilfe
  Themen

- Klicken Sie im Fenster “Hilfethemen” auf die Registerkarte Index.

- Geben Sie Text-Viewer ein, und doppelklicken Sie auf den entsprechenden Indexeintrag.

Ein- und Ausblenden von Ergebnissen

Im Viewer können Sie ausgewählte Tabellen oder alle Ergebnisse einer Prozedur ein- und ausblenden. Dies ist nützlich, wenn Sie möchten, daß im Inhaltsfenster weniger angezeigt wird.

So blenden Sie Tabellen und Diagramme aus

- Doppelklicken Sie im Gliederungsfenster des Viewers auf das Buchsymbol des Objekts.

  oder

- Klicken Sie einmal auf das Objekt, um es auszuwählen.

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Ansicht
  Ausblenden

  oder

- Klicken Sie in der Gliederungs-Symbolleiste auf die Schaltfläche “Ausblenden”, die durch ein geschlossenes Buch dargestellt ist.

Die Schaltfläche “Einblenden”, die durch ein geöffnetes Buch dargestellt ist, wird zu einer aktiven Schaltfläche. Dadurch wird angezeigt, daß das Objekt jetzt ausgeblendet ist.

So blenden Sie die Ergebnisse einer Prozedur aus

- Klicken Sie im Gliederungsfenster auf das Kästchen neben dem Namen der Prozedur.

  Dadurch werden alle Ergebnisse der Prozedur ausgeblendet und die Gliederungsansicht reduziert.

Verschieben, Löschen und Kopieren von Ausgaben

Sie können die Ergebnisse neu anordnen, indem Sie einzelne Objekte oder ganze Objektgruppen kopieren, verschieben oder löschen.

So verschieben Sie Ausgaben im Viewer

- Klicken Sie auf ein Objekt im Gliederungs- oder Inhaltsfenster. Sie können mehrere aufeinanderfolgende Objekte auswählen, indem Sie die Umschalttaste gedrückt halten und auf das erste und das letzte Objekt klicken. Bei nicht aufeinanderfolgenden Objekten klicken Sie auf die einzelnen Objekte und halten dabei die Strg-Taste gedrückt.
- Klicken Sie die ausgewählten Objekte, und ziehen Sie sie bei gedrückter Maustaste.
- Ziehen Sie die Objekte mit gedrückter Maustaste. Legen Sie die Objekte an der gewünschten Stelle ab, indem Sie die Maustaste loslassen.

  Sie können Objekte auch mit den Befehlen “Ausschneiden” und “Einfügen nach” aus dem Menü “Bearbeiten” verschieben.

So löschen Sie Ausgaben im Viewer

- Klicken Sie auf ein Objekt im Gliederungs- oder Inhaltsfenster. Sie können mehrere aufeinanderfolgende Objekte auswählen, indem Sie die Umschalttaste gedrückt halten und auf das erste und das letzte Objekt klicken. Bei nicht aufeinanderfolgenden Objekten klicken Sie auf die einzelnen Objekte und halten dabei die Strg-Taste gedrückt.

- Drücken Sie die Entf-Taste.

  oder

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Löschen

So kopieren Sie Ausgaben im Viewer

- Klicken Sie auf ein Objekt im Gliederungs- oder Inhaltsfenster. Sie können mehrere aufeinanderfolgende Objekte auswählen, indem Sie die Umschalttaste gedrückt halten und auf das erste und das letzte Objekt klicken. Bei nicht aufeinanderfolgenden Objekten klicken Sie auf die einzelnen Objekte und halten dabei die Strg-Taste gedrückt.
- Halten Sie die Strg-Taste beim Klicken und Ziehen der Objekte gedrückt. (Halten Sie für das Ziehen die Maustaste gedrückt.)
- Legen Sie die Objekte an der gewünschten Stelle ab, indem Sie die Maustaste loslassen.

  Sie können Objekte auch mit den Befehlen “Kopieren” und “Einfügen nach” aus dem Menü “Bearbeiten” kopieren.

Ändern der anfänglichen Ausrichtung

In der Standardeinstellung sind alle Ergebnisse linksbündig ausgerichtet. Sie können die anfängliche Ausrichtung ändern. Wählen Sie dazu im Menü “Bearbeiten” den Befehl Optionen, und klicken Sie anschließend auf die Registerkarte Viewer.

Ändern der Ausrichtung von Ausgabeobjekten

- Klicken Sie im Gliederungs- oder Inhaltsfenster auf die Objekte, die Sie ausrichten möchten. (Mit gedrückter Umschalt- oder Strg-Taste können Sie mehrere Objekte auswählen.)
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Format
  Linksbündig

Ihnen stehen außerdem die Optionen “Zentriert” und “Rechtsbündig” zur Verfügung.

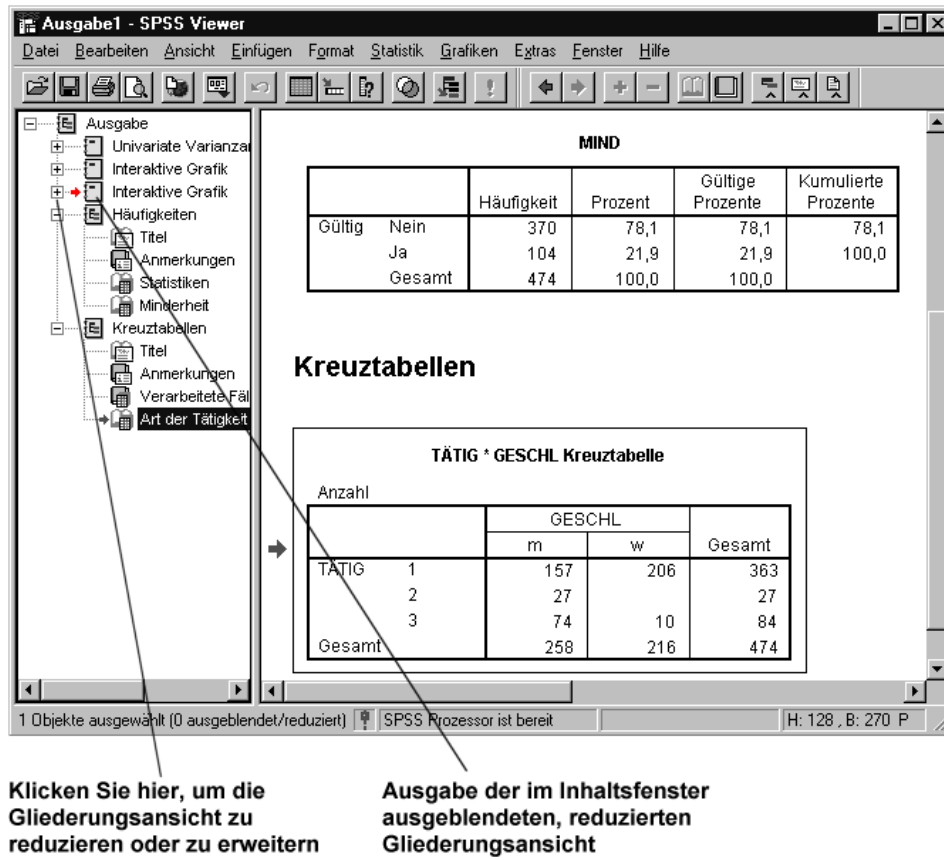
Anmerkung: Im Viewer werden alle Ergebnisse linksbündig angezeigt. Die Einstellungen für die Ausrichtung wirken sich nur auf gedruckte Ausgaben aus. Zentrierte und rechtsbündig ausgerichtete Objekte werden durch kleine Symbole an der linken oberen Ecke des Objekts gekennzeichnet.

Gliederung des Viewers

Im Gliederungsfenster wird eine Inhaltsangabe des Viewer-Dokuments angezeigt. Hier können Sie die Ergebnisse durchblättern und festlegen, welche Objekte angezeigt werden. Die meisten Aktionen im Gliederungsfenster wirken sich auch auf das Inhaltsfenster aus.

- Beim Auswählen eines Objekts im Gliederungsfenster wird das entsprechende Objekt im Inhaltsfenster angezeigt.
- Beim Verschieben eines Objekts im Gliederungsfenster wird das entsprechende Objekt im Inhaltsfenster ebenfalls verschoben.
- Beim Reduzieren der Gliederungsansicht werden die Ergebnisse aller Objekte in den reduzierten Ebenen ausgeblendet.

Abbildung 10-2
Reduzierte Gliederungsansicht und ausgeblendete Ergebnisse



Einstellen der Anzeige für die Gliederung. Zum Einstellen der Anzeige für die Gliederung stehen Ihnen die folgenden Möglichkeiten zur Verfügung:

- Erweitern und Reduzieren der Gliederungsansicht,
- Ändern der Gliederungsebene von ausgewählten Objekten,
- Ändern der Größe von Objekten im Gliederungsfenster,
- Ändern der in der Gliederung verwendeten Schriftart.

So reduzieren und erweitern Sie die Gliederungsansicht

- Klicken Sie auf das Kästchen links neben dem Gliederungsobjekt, das Sie reduzieren oder erweitern möchten.

  oder

- Klicken Sie in der Gliederung auf das Objekt.

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Ansicht
  Reduzieren

  oder

  Ansicht
  Erweitern

So ändern Sie die Gliederungsebene

- Klicken Sie im Gliederungsfenster auf das Objekt.

- Klicken Sie in der Gliederungs-Symbolleiste auf den nach links zeigenden Pfeil, um das Objekt heraufzustufen (Verschieben des Objekts nach links).

- Klicken Sie in der Gliederungs-Symbolleiste auf den nach rechts zeigenden Pfeil, um das Objekt herabzustufen (Verschieben des Objekts nach rechts).

  oder

  Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Gliederung
  Heraufstufen

  oder

  Bearbeiten
  Gliederung
  Herabstufen

  Das Ändern der Gliederungsebene ist besonders nach dem Verschieben von Objekten im Gliederungsfenster nützlich. Beim Verschieben der Objekte kann sich die Gliederungsebene der Objekte ändern. In diesem Fall können Sie die Pfeil-Schaltflächen auf der Gliederungs-Symbolleiste verwenden, um die ursprüngliche Gliederungsebene wiederherzustellen.

So ändern Sie die Größe von Objekten in der Gliederung

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Ansicht
  Größe der Gliederung
    Klein

Es stehen außerdem die Optionen “Mittel” und “Groß” zur Verfügung.

Die Größe der Symbole und der dazugehörigen Texte wird geändert.

So ändern Sie die Schriftart in der Gliederung

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Ansicht
  Schriftart für Gliederung...

▶ Wählen Sie eine Schriftart aus.

Einfügen von Objekten im Viewer

Sie können im Viewer Objekte einfügen, beispielsweise Titel, neue Texte, Diagramme oder Objekte aus anderen Anwendungen.

So fügen Sie einen neuenTitel oder ein neues Textobjekt ein

Sie können im Viewer Textobjekte einfügen, die nicht mit einer Tabelle oder einem Diagramm verbunden sind.

▶ Klicken Sie auf die Tabelle, das Diagramm oder das Objekt, nach der bzw. dem der Titel oder Text eingefügt werden soll.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Einfügen
  Neuer Titel

oder

Einfügen
  Neuer Text

▶ Doppelklicken Sie auf das neue Objekt.

- Geben Sie den Text ein, der an dieser Stelle stehen soll.

So fügen Sie eine Textdatei hinzu

- Klicken Sie im Gliederungs- oder Inhaltsfenster des Viewers auf die Tabelle, das Diagramm oder das Objekt, nach der bzw. dem der Text eingefügt werden soll.
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Einfügen
  Textdatei...
- Wählen Sie eine Textdatei aus.

  Doppelklicken Sie auf die Textdatei, um diese zu bearbeiten.

Verwenden von Ausgaben in anderen Anwendungen

Sie können Pivot-Tabellen und Diagramme in anderen Windows-Anwendungen verwenden, zum Beispiel in Textverarbeitungs- oder Tabellenkalkulationsprogrammen. Die Pivot-Tabellen oder Diagramme können dabei in verschiedenen Formaten eingefügt werden, beispielsweise als:

Eingebettetes Objekt. In Anwendungen, die ActiveX-Objekte unterstützen, können Sie Pivot-Tabellen und interaktive Diagramme einbetten. Nach dem Einfügen der Tabelle kann diese direkt in der Anwendung durch Doppelklicken aktiviert werden. Anschließend können Sie die Tabelle wie im Viewer bearbeiten.

Grafik (Metadatei). Sie können Pivot-Tabellen, Textausgaben und Diagramme als Grafiken im Metadatei-Format einfügen. Die Größe der Grafiken kann in anderen Anwendungen geändert werden, und unter Umständen können die Grafiken in begrenztem Umfang auch mit den Funktionen der anderen Anwendungen bearbeitet werden. Bei Pivot-Tabellen, die als Bilder eingefügt wurden, bleiben alle Rahmen und Schriftartenmerkmale erhalten.

RTF (Rich Text Format). Pivot-Tabellen können in andere Anwendungen im RTF-Format eingefügt werden. In den meisten Anwendungen wird die Pivot-Tabelle dabei als Tabelle eingefügt, die dann in der anderen Anwendung bearbeitet werden kann.

Bitmap. Diagramme können in andere Anwendungen als Bitmaps eingefügt werden.

BIFF. Der Inhalt von Tabellen kann in eine Tabellenkalkulation eingefügt werden, wobei die numerische Genauigkeit erhalten bleibt.

Text. Der Inhalt von Tabellen kann als Text kopiert und in andere Anwendungen eingefügt werden. Dies kann bei Anwendungen wie E-Mail-Programmen nützlich sein, bei denen mit der Anwendung nur Text verarbeitet oder übertragen werden kann.

So kopieren Sie Tabellen und Diagramme

- Wählen Sie die zu kopierende Tabelle oder das Diagramm aus.
- Wählen Sie die folgenden Befehle aus den Menüs aus:

Bearbeiten
  Kopieren

So kopieren Sie Ergebnisse und fügen diese in andere Anwendungen ein

- Kopieren Sie die Ergebnisse im Viewer.
- Wählen Sie die folgenden Befehle aus den Menüs der Ziel-Anwendung aus:

Bearbeiten
  Einfügen

oder

Bearbeiten
  Inhalte einfügen...

Einfügen. Ausgaben werden in einer Reihe von Formaten in die Zwischenablage kopiert. Jede Anwendung ermittelt das "beste" Format für das Einfügen. In vielen Anwendungen werden die Ergebnisse mit dem Befehl "Einfügen" als Bild (Metadatei) eingefügt. In Textverarbeitungsprogrammen werden Pivot-Tabellen mit dem Befehl "Einfügen" im RTF-Format in Form einer Tabelle eingefügt. Bei Tabellenkalkulationsprogrammen werden Pivot-Tabellen mit dem Befehl "Einfügen" im BIFF-Format eingefügt. Diagramme werden als Metadateien eingefügt.

Inhalte einfügen. Ergebnisse werden in einer Reihe von Formaten in die Zwischenablage kopiert. Mit dem Befehl "Inhalte einfügen" können Sie das gewünschte Format aus der Liste der verfügbaren Formate in der Ziel-Anwendung auswählen.

So betten Sie eine Tabelle in andere Anwendungen ein

Sie können Pivot-Tabellen und interaktive Diagramme im ActiveX-Format in andere Anwendungen einbetten. Ein eingebettetes Objekt kann direkt in der Anwendung durch Doppelklicken aktiviert und anschließend wie im Viewer bearbeitet und pivotiert werden.

Wenn Sie mit Anwendungen arbeiten, die ActiveX-Objekte unterstützen:

- Doppelklicken Sie auf die Datei *objs-on.bat*. Diese Datei befindet sich in dem Verzeichnis, in dem SPSS installiert wurde.

  Durch diesen Vorgang wird die ActiveX-Einbettung für Pivot-Tabellen aktiviert. Mit der Datei *objs-off.bat* wird die ActiveX-Einbettung wieder deaktiviert.

So betten Sie eine Pivot-Tabelle oder ein interaktives Diagramm in eine andere Anwendung ein:

- Kopieren Sie die Tabelle im Viewer.
- Wählen Sie die folgenden Befehle aus den Menüs der Ziel-Anwendung aus:

  Bearbeiten
  Inhalte einfügen...

- Wählen Sie aus der Liste die Option Objekt: SPSS Pivot-Tabelle*oder*SPSS Interactive Graph aus.

  Die Ziel-Anwendung muß ActiveX-Objekte unterstützen. Informationen zur ActiveX-Unterstützung finden Sie im Handbuch der jeweiligen Anwendung. Bei einigen Anwendungen, die ActiveX nicht unterstützen, können möglicherweise trotzdem ActiveX-Pivot-Tabellen eingefügt werden. Sie müssen aber in diesem Fall mit einem instabilen Verhalten der Anwendung rechnen. Verlassen Sie sich nicht auf eingebettete Objekte, bis Sie die Stabilität der Anwendung mit eingebetteten ActiveX-Objekten getestet haben.

So fügen Sie Pivot-Tabellen oder Diagramme als Bilder (Metadateien) ein

- Kopieren Sie die Tabelle oder das Diagramm im Viewer.
- Wählen Sie die folgenden Befehle aus den Menüs der Ziel-Anwendung aus:

  Bearbeiten
  Inhalte einfügen...

- Wählen Sie als Format Bild aus der Liste aus.

  Das Objekt wird als Metadatei eingefügt. Es werden nur die Schichten und Spalten der Tabelle eingefügt, die beim Kopieren des Objekts sichtbar waren. Andere Schichten und ausgeblendete Spalten sind nicht verfügbar.

So fügen Sie Pivot-Tabellen als Tabellen (RTF) ein

- Kopieren Sie die Pivot-Tabelle im Viewer.
- Wählen Sie die folgenden Befehle aus den Menüs der Ziel-Anwendung aus:

  Bearbeiten
  Inhalte einfügen...

- Wählen Sie Formatierten Text (RTF) oder Rich Text Format aus der Liste aus.

  Die Pivot-Tabelle wird als Tabelle eingefügt. Es werden nur die Schichten und Spalten in die Tabelle eingefügt, die beim Kopieren des Objekts sichtbar waren. Andere Schichten und ausgeblendete Spalten sind nicht verfügbar. Sie können nicht mehrere Pivot-Tabellen gleichzeitig in diesem Format kopieren und einfügen.

So fügen Sie Pivot-Tabellen als Text ein

- Kopieren Sie die Tabelle im Viewer.
- Wählen Sie die folgenden Befehle aus den Menüs der Ziel-Anwendung aus:

  Bearbeiten
  Inhalte einfügen...

- Wählen Sie Unformatierten Text aus der Liste aus.

  Bei unformatiertem Text aus Pivot-Tabellen sind die einzelnen Spalten durch Tabulatoren getrennt. Sie können Spalten ausrichten, indem Sie in der anderen Anwendung die Tabstops entsprechend anpassen.

So kopieren Sie mehrere Objekte und fügen diese in eine andere Anwendung ein

- Wählen Sie die zu kopierenden Tabellen und/oder Diagramme aus. Halten Sie zum Auswählen mehrerer Objekte beim Klicken mit der Maus gleichzeitig die Umschalt- oder die Strg-Taste gedrückt.

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Objekte kopieren

- Wählen Sie die folgenden Befehle aus den Menüs der Ziel-Anwendung aus:

  Bearbeiten
  Einfügen

  Anmerkung: Verwenden Sie den Befehl “Objekte kopieren” nur zum Kopieren mehrerer Objekte aus dem Viewer in andere Anwendungen. Verwenden Sie zum Kopieren und Einfügen in Viewer-Dokumenten, beispielsweise zwischen zwei Viewer-Fenstern, den Befehl “Kopieren” im Menü “Bearbeiten”.

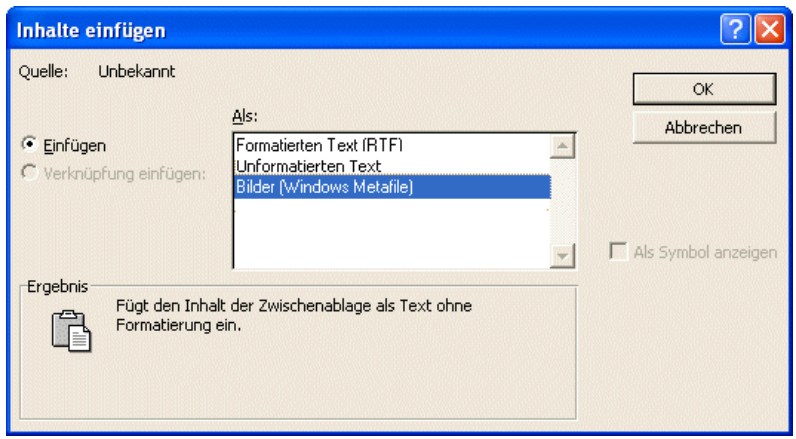
Einfügen von Objekten im Viewer

Sie können Objekte aus anderen Anwendungen in den Viewer einfügen. Verwenden Sie hierfür den Befehl “Einfügen nach” oder “Inhalte einfügen”. Dabei wird das neue Objekt nach dem aktuell ausgewählten Objekt im Viewer eingefügt. Verwenden Sie den Befehl “Inhalte einfügen”, wenn Sie das Format für das einzufügende Objekt auswählen möchten.

Inhalte einfügen

Mit dem Befehl “Inhalte einfügen” können Sie das Format von kopierten Objekten auswählen, die im Viewer eingefügt werden. Die möglichen Dateitypen für das Objekt werden im Dialogfeld “Inhalte einfügen” aufgeführt. Das Objekt wird im Viewer nach dem gegenwärtig ausgewählten Objekt eingefügt.

Abbildung 10-3
Dialogfeld "Inhalte einfügen"



Einfügen von Objekten aus anderen Anwendungen in den Viewer

- Kopieren Sie das Objekt in der anderen Anwendung.
- Klicken Sie im Gliederungs- oder Inhaltsfenster des Viewers auf die Tabelle, das Diagramm oder das Objekt, nach der bzw. dem das Objekt eingefügt werden soll.
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Inhalte einfügen...
- Wählen Sie das Format für das Objekt aus der Liste aus.

Ausgabe exportieren

Im Dialogfeld "Ausgabe exportieren" können Sie festlegen, dass Pivot-Tabellen und Textausgaben im HTML-, Text-, Word-/RTF-, Excel- oder PowerPoint-Format (erfordert PowerPoint 97 oder höher) gespeichert werden. Sie können auch festlegen, daß Diagramme in verschiedenen gebräuchlichen Formaten gespeichert werden, die auch von anderen Anwendungen verwendet werden. (*Anmerkung*: Der Export nach PowerPoint ist in der Studentenversion nicht verfügbar.)

Ausgabedokument. Hiermit können Sie beliebige Kombinationen von Pivot-Tabellen, Textausgaben und Diagrammen exportieren.

- Beim HTML- und Textformat werden Diagramme im derzeit ausgewählten Exportformat für Diagramme exportiert. Bei Dokumenten im HTML-Format werden Diagramme als Verweis eingebettet. Exportieren Sie Diagramme in diesem Fall in einem für die Aufnahme in HTML-Dokumente geeigneten Format. Bei Dokumenten im Textformat wird in der Textdatei für jedes Diagramm eine Zeile mit der Angabe des Dateinamens für das exportierte Diagramm eingefügt.
- Beim Word/RTF-Format werden Diagramme im Windows-Metafile-Format exportiert und in das Word-Dokument eingebettet.
- In Excel-Dokumente können keine Diagramme exportiert werden.
- Beim PowerPoint-Format werden Diagramme im TIFF-Format exportiert und in die PowerPoint-Datei eingebettet.

Ausgabedokument (ohne Diagramme). Hiermit werden Pivot-Tabellen und Textausgaben exportiert. Sämtliche Diagramme im Viewer werden ignoriert.

Nur Diagramme. Folgende Exportformate sind verfügbar: Windows-Metadateien (WMF), Enhanced Metafile (EMF), Windows-Bitmaps (BMP), Encapsulated PostScript (EPS), JPEG, TIFF, PNG und Macintosh-PICT.

Was exportieren. Sie können alle Objekte im Viewer, alle sichtbaren Objekte oder nur ausgewählte Objekte exportieren.

Exportformat. Bei Ausgabedokumenten haben Sie die Wahl zwischen HTML-, Text-, Excel-, Word-/RTF- und PowerPoint-Format. Bei HTML- und Textformat werden Diagramme in dem Format exportiert, das gegenwärtig im Dialogfeld “Optionen” für das ausgewählte Format festgelegt ist. Wenn Sie die Option “Nur Diagramme” ausgewählt haben, wählen Sie ein Exportformat aus der Dropdown-Liste aus. Bei Ausgabedokumenten werden Pivot-Tabellen und Text folgendermaßen exportiert:

- **HTML-Datei (*.htm).** Pivot-Tabellen werden als HTML-Tabellen exportiert. Textausgaben werden als vorformatierter HTML-Text exportiert.
- **Textdatei (*.txt).** Pivot-Tabellen können als durch Tabulatoren getrennter Text oder als durch Leerzeichen getrennter Text exportiert werden. Alle Textausgaben werden in durch Leerzeichen getrenntem Format exportiert.
- **Excel-Datei (*.xls).** Die Zeilen, Spalten und Zellen von Pivot-Tabellen werden mit sämtlichen Formatierungsattributen wie Zellenrahmen, Schriftarten, Hintergrundfarben als Excel-Zeilen, -Spalten und -Zellen exportiert.

Textausgaben werden mit allen Schriftartattributen exportiert. Jede Zeile in der Textausgabe entspricht einer Zeile in der Excel-Datei, wobei der gesamte Inhalt der Zeile in einer einzelnen Zelle enthalten ist.

- **Word/RTF-Datei (*.doc).** Pivot-Tabellen werden mit sämtlichen Formatierungsattributen wie Zellenrahmen, Schriftarten, Hintergrundfarben usw. als Word-Tabellen exportiert. Textausgaben werden als formatierter RTF-Text exportiert. Textausgaben in SPSS werden immer mit einem nicht proportionalen Zeichensatz (mit festem Abstand) angezeigt und mit denselben Schriftartenattributen exportiert. Für die richtige Ausrichtung von durch Leerzeichen getrennten Textausgaben ist ein nicht proportionaler Zeichensatz (mit festem Abstand) erforderlich.
- **PowerPoint file (*.ppt).** Pivot-Tabellen werden als Word-Dateien exportiert und sind auf separaten Folien in der PowerPoint-Datei eingebettet (je eine Pivot-Tabelle auf einer Folie). Sämtliche Formatierungsattribute der Pivot-Tabelle (z. B. Zellenrahmen, Schriftarten und Hintergrundfarben) werden beibehalten. Textausgaben werden als formatierter RTF-Text exportiert. Textausgaben in SPSS werden immer mit einem nicht proportionalen Zeichensatz (mit festem Abstand) angezeigt und mit denselben Schriftartenattributen exportiert. Für die richtige Ausrichtung von durch Leerzeichen getrennten Textausgaben ist ein nicht proportionaler Zeichensatz (mit festem Abstand) erforderlich. (*Anmerkung*: Der Export nach PowerPoint ist in der Studentenversion nicht verfügbar.)

Ausgabeverwaltungssystem. Sie können auch automatisch alle Ausgaben oder vom Benutzer festgelegte Ausgabetypen als Datendateien im Text-, HTML-, XML- oder SPSS-Format exportieren. Für weitere Informationen siehe “Ausgabeverwaltungssystem (OMS)” in Kapitel 48 auf S. 747.

So exportieren Sie Ausgaben

- ▶ Wechseln Sie in das Viewer-Fenster. (Klicken Sie auf eine beliebige Stelle im Fenster.)
- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

  Datei
  Exportieren...

- ▶ Geben Sie einen Dateinamen oder ein Präfix für Diagramme ein, und wählen Sie ein Exportformat aus.

Abbildung 10-4
Dialogfeld "Ausgabe exportieren"

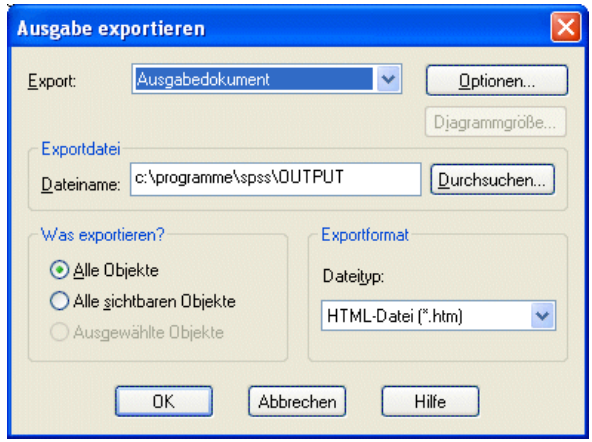
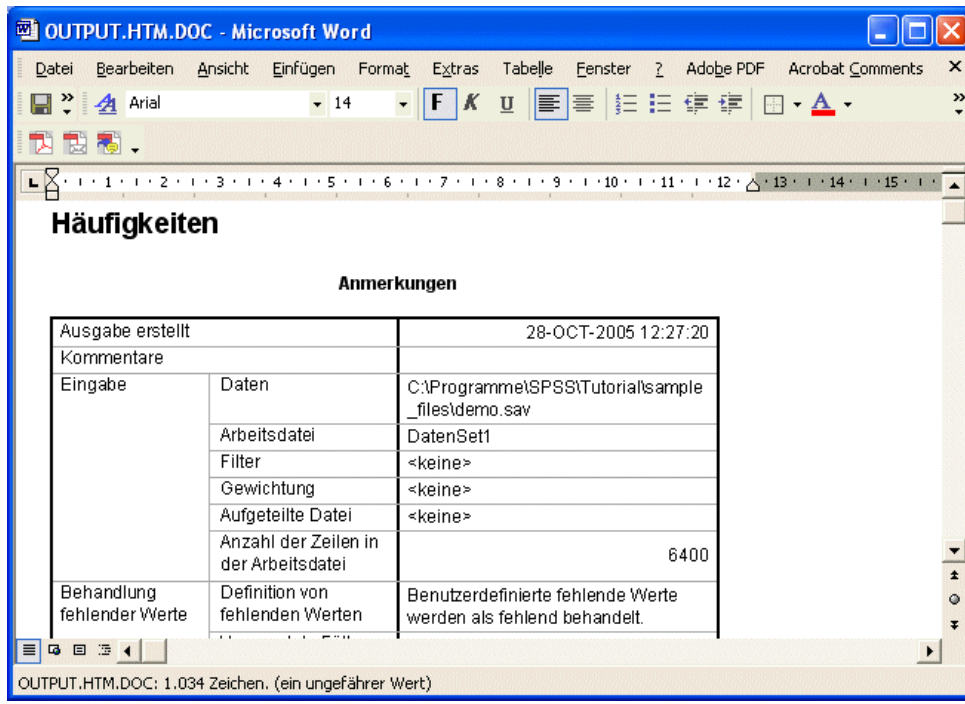


Abbildung 10-5
Im Word/RTF-Format exportierte Ausgabe



Optionen für HTML, Word/RTF und Excel

Wenn Sie im Dialogfeld “Ausgabe exportieren” auf “Optionen” klicken, wird ein weiteres Dialogfeld angezeigt. Hier wird festgelegt, ob Fußnoten und Erklärungen in Dokumente eingeschlossen werden, die in das HTML-, Word-/RTF- oder Excel-Format exportiert werden. Außerdem werden die Diagrammexport-Optionen für HTML-Dokumente und die Behandlung von mehrschichtigen Pivot-Tabellen festgelegt.

Bildformat. Hier werden das Exportformat für Diagramme und optionale Einstellungen, z. B. die Diagrammgröße für HTML-Dokumente, festgelegt. Im Falle des Word/RTF-Formats werden alle Diagramme im Windows-Metadatei-Format (WMF) exportiert. Bei Excel werden keine Diagramme eingeschlossen.

Fußnoten und Erklärungen exportieren. Mit diesem Kontrollkästchen geben Sie an, daß alle Fußnoten und Erklärungen beim Export der Pivot-Tabellen berücksichtigt werden sollen.

Alle Schichten exportieren. Aktivieren Sie dieses Kontrollkästchen, um alle Schichten einer mehrschichtigen Pivot-Tabelle zu exportieren. Wenn dieses Kontrollkästchen deaktiviert verbleibt, wird lediglich die oberste Schicht exportiert.

So legen Sie die Exportoptionen für HTML, Word/RTF und Excel fest

- Wechseln Sie in das Viewer-Fenster. (Klicken Sie auf eine beliebige Stelle im Fenster.)
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Datei
  Exportieren...

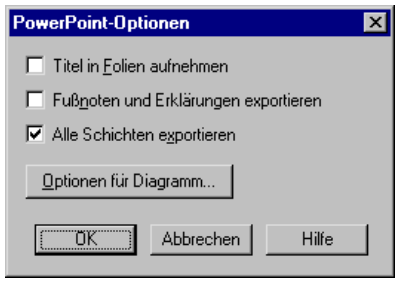
- Wählen Sie als Exportformat HTML-Datei, Word/RTF-Datei oder Excel-Datei aus.
- Klicken Sie auf Optionen.

PowerPoint-Optionen

Die PowerPoint-Optionen steuern das Einschließen von Titeln für Folien, die Aufnahme von Fußnoten und Erklärungen für Pivot-Tabellen, die Behandlung mehrschichtiger Pivot-Tabellen sowie die Optionen für Diagramme, die nach

PowerPoint exportiert werden. (*Anmerkung*: Der Export nach PowerPoint ist in der Studentenversion nicht verfügbar.)

Abbildung 10-6
Dialogfeld "PowerPoint-Optionen"



Titel in Dias aufnehmen. Aktivieren Sie dieses Kontrollkästchen, damit auf jeder beim Export erzeugten Folie ein Titel eingefügt wird. Jede Folie enthält ein einzelnes Element, das aus dem Viewer exportiert wurde. Der Titel wird aus dem Gliederungseintrag für das Element im Gliederungsfenster des Viewers gebildet.

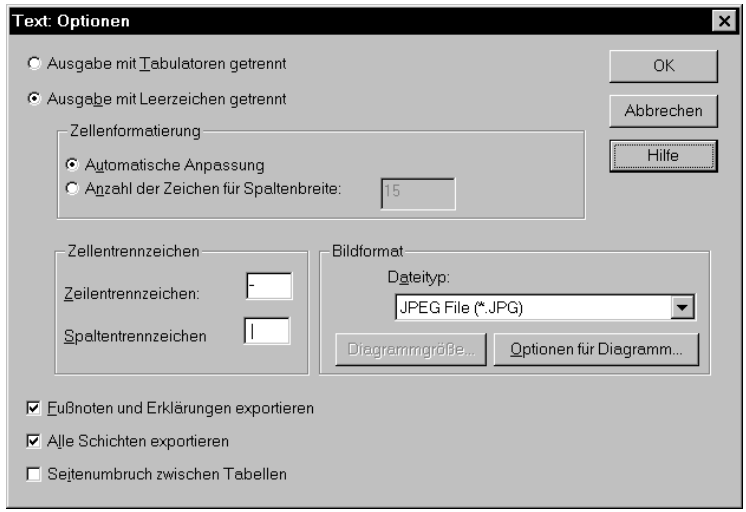
Fußnoten und Erklärungen exportieren. Mit diesem Kontrollkästchen geben Sie an, daß alle Fußnoten und Erklärungen beim Export der Pivot-Tabellen berücksichtigt werden sollen.

Alle Schichten exportieren. Aktivieren Sie dieses Kontrollkästchen, damit alle Schichten einer mehrschichtigen Pivot-Tabelle exportiert werden. Jede Schicht wird dabei auf eine separate Folie platziert, und alle Schichten erhalten denselben Titel. Wenn dieses Kontrollkästchen deaktiviert verbleibt, wird lediglich die oberste Schicht exportiert.

Text: Optionen

Im Dialogfeld "Text: Optionen" können Sie die Formatoptionen für Pivot-Tabellen, Textausgaben und Diagramme sowie die Aufnahme von Fußnoten und Erklärungen für Dokumente festlegen, die im Textformat exportiert werden.

Abbildung 10-7
Dialogfeld "Text: Optionen"



Pivot-Tabellen können als durch Tabulatoren getrennter Text oder als durch Leerzeichen getrennter Text exportiert werden. Wenn eine Zelle nicht leer ist, wird beim durch Tabulatoren getrennten Textformat der Inhalt der Zelle und ein Tabulatorzeichen ausgegeben. Wenn eine Zelle leer ist, wird ein Tabulatorzeichen ausgegeben.

Alle Textausgaben werden in durch Leerzeichen getrenntem Format exportiert. Zur richtigen Ausrichtung von durch Leerzeichen getrennten Textausgaben ist ein Zeichensatz mit festem Abstand nötig.

Zellenformatierung. Bei mit Leerzeichen getrennten Pivot-Tabellen werden in der Standardeinstellung alle Zeilenumbrüche entfernt und jede Spalte nach der Breite der längsten Beschriftung oder des längsten Werts in der Spalte eingerichtet. Wenn Sie die Spaltenbreite beschränken und lange Beschriftungen umbrechen möchten, geben Sie eine Anzahl von Zeichen für die Spaltenbreite an. Diese Einstellung betrifft nur Pivot-Tabellen.

Zellentrennzeichen. Bei mit Leerzeichen getrennten Pivot-Tabellen können Sie die Zeichen festlegen, die zum Trennen der Zellen verwendet werden.

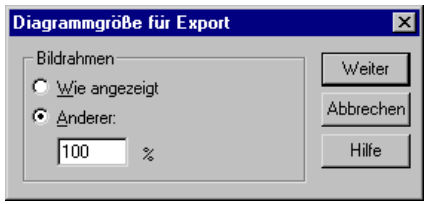
Bildformat. Hier werden das Exportformat für Diagramme und optionale Einstellungen, z. B. die Diagrammgröße, festgelegt.

Seitenumbruch zwischen Tabellen. Fügt zwischen den einzelnen Tabellen einen Seitenvorschub/-umbruch ein. Mithilfe dieser Option wird bei mehrschichtigen Pivot-Tabellen ein Seitenumbruch zwischen den einzelnen Schichten eingefügt.

Optionen für die Diagrammgröße

Durch die Einstellung der Diagrammgröße wird die Größe der exportierten Diagramme festgelegt. Die Prozentangabe ermöglicht eine Verkleinerung oder Vergrößerung des exportierten Diagramms um bis zu 200 %.

Abbildung 10-8
Dialogfeld "Diagrammgröße für Export"



So legen Sie die Größe für exportierte Diagramme fest

- Wechseln Sie in das Viewer-Fenster. (Klicken Sie auf eine beliebige Stelle im Fenster.)
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Datei
  Exportieren...
- Klicken Sie bei Wahl der Option "Ausgabedokument" auf Optionen, und wählen Sie das Exportformat. Klicken Sie dann auf Diagrammgröße.
- Wenn Sie die Option "Nur Diagramme" ausgewählt haben, wählen Sie das Exportformat, und klicken Sie dann auf Diagrammgröße.

Exportoptionen für JPEG-Diagramme

Farbtiefe. JPEG-Diagramme können mit der Einstellung True Color (24-Bit) oder 256 Graustufen exportiert werden.

Farbraum. Die Option "Farbraum" steuert die Art der Farbkodierung im Bild. Das YUV-Farbmodell ist eine Form der Farbkodierung, die bei digitalen Video- und MPEG-Übertragungen sehr gebräuchlich ist. YUV ist eine Abkürzung für Y-Signal, U-Signal und V-Signal. Durch die Y-Komponente wird der Graustufenwert oder die Luminanz angegeben. Die U- und V-Komponenten entsprechen der Chrominanz (Farbinformation).

Die Verhältnisse dieser Werte zueinander entsprechen den Abtastraten für jede Komponente. Durch eine Reduzierung der U- und V-Abtastraten verringert sich die Dateigröße (und die Qualität). Durch den Farbraum wird der Grad der Farbverluste im exportierten Bild angegeben. YUV 4:4:4 ist verlustfrei, während YUV 4:2:2 und YUV 4:1:1 bei einer geringeren Dateigröße (Speicherplatz) auch eine geringere Bildqualität bietet.

Progressive Kodierung. Ermöglicht das stufenweise Laden von Bildern, die anfänglich mit einer geringen Auflösung angezeigt und dann während des weiteren Ladevorgangs qualitativ verbessert werden.

Einstellung der Komprimierungsqualität. Steuert das Verhältnis der Komprimierung zur Bildqualität. Je besser die Bildqualität, desto umfangreicher ist die Größe der exportierten Datei.

Farboperationen. Die folgenden Operationen sind verfügbar:

- **Invertieren.** Die einzelnen Pixel werden entgegengesetzt ihrer ursprünglichen Farbeinstellung gespeichert.
- **Gamma-Korrektur.** Paßt die Farbintensität im exportierten Diagramm durch Änderung der Gamma-Konstante an, mit der Intensitätswerte zugeordnet werden. Mithilfe dieser Option kann das Bitmap-Bild aufgehellt oder abgedunkelt werden. Die Werte können im Bereich von 0,10 (ganz dunkel) bis 6,5 (ganz hell) liegen.

Exportoptionen für BMP- und PICT-Diagramme

Farbtiefe. Bestimmt die Anzahl der Farben im exportierten Diagramm. Ein Diagramm, das mit einer beliebigen Farbtiefe gespeichert wird, verfügt über eine Mindestzahl tatsächlich verwendeter Farben und eine Höchstzahl zulässiger Farben in dieser Farbtiefe. Wenn das Diagramm beispielsweise die drei Farben Rot, Weiß und Schwarz enthält, jedoch als Bild mit 16 Farben gespeichert wird, enthält das Diagramm weiterhin lediglich drei Farben.

- Wenn die Anzahl der Farben im Diagramm die Anzahl der Farben für diese Tiefe übersteigt, werden die Farben zur Replizierung der Diagrammfarben gemischt.
- Aktuelle Bildschirmtiefe ist die Anzahl der Farben, die gegenwärtig auf Ihrem Monitor dargestellt werden.

Farboperationen. Die folgenden Operationen sind verfügbar:

- **Invertieren.** Die einzelnen Pixel werden entgegengesetzt ihrer ursprünglichen Farbeinstellung gespeichert.
- **Gamma-Korrektur.** Paßt die Farbintensität im exportierten Diagramm durch Änderung der Gamma-Konstante an, mit der Intensitätswerte zugeordnet werden. Mithilfe dieser Option kann das Bitmap-Bild aufgehellt oder abgedunkelt werden. Die Werte können im Bereich von 0,10 (ganz dunkel) bis 6,5 (ganz hell) liegen.

RLE-Komprimierung verwenden. (nur BMP) Ein verlustfreies Komprimierungsverfahren, das von den gebräuchlichsten Windows-Dateiformaten unterstützt wird. Verlustfreie Komprimierung bedeutet, daß die Dateigröße ohne Beeinträchtigung der Bildqualität reduziert wird.

Exportoptionen für PNG- und TIFF-Diagramme

Farbtiefe. Bestimmt die Anzahl der Farben im exportierten Diagramm. Ein Diagramm, das mit einer beliebigen Farbtiefe gespeichert wird, verfügt über eine Mindestzahl tatsächlich verwendeter Farben und eine Höchstzahl zulässiger Farben in dieser Farbtiefe. Wenn das Diagramm beispielsweise die drei Farben Rot, Weiß und Schwarz enthält, jedoch als Bild mit 16 Farben gespeichert wird, enthält das Diagramm weiterhin lediglich drei Farben.

- Wenn die Anzahl der Farben im Diagramm die Anzahl der Farben für diese Tiefe übersteigt, werden die Farben zur Replizierung der Diagrammfarben gemischt.
- Aktuelle Bildschirmtiefe ist die Anzahl der Farben, die gegenwärtig auf Ihrem Monitor dargestellt werden.

Farboperationen. Die folgenden Operationen sind verfügbar:

- **Invertieren.** Die einzelnen Pixel werden entgegengesetzt ihrer ursprünglichen Farbeinstellung gespeichert.
- **Gamma-Korrektur.** Paßt die Farbintensität im exportierten Diagramm durch Änderung der Gamma-Konstante an, mit der Intensitätswerte zugeordnet werden. Mithilfe dieser Option kann das Bitmap-Bild aufgehellt oder abgedunkelt werden. Die Werte können im Bereich von 0,10 (ganz dunkel) bis 6,5 (ganz hell) liegen.

Transparenz. Ermöglicht Ihnen die Auswahl einer Farbe, die im exportierten Diagramm transparent erscheint. Diese Funktion ist nur beim Export mit True Color (32-Bit) verfügbar. Geben Sie für jede Farbe einen ganzzahligen Wert von 0 bis 255 ein. Der Standardwert für jede Farbe ist 255. Dieser Wert steht für die transparente Standardfarbe Weiß.

Format. (Nur für TIFF) Mit dieser Option können Sie den Farbraum festlegen und das exportierte Diagramm komprimieren. Für RGB-Farben sind alle Farbtiefen verfügbar. Für CMYK-Farben kann nur True Color (24-Bit) und True Color (32-Bit) ausgewählt werden. Bei Wahl der Option "YCbCr" ist nur True Color (24-Bit) verfügbar.

Exportoptionen für EPS-Diagramme

Bäume, Karten und interaktive Diagramme

Für Bäume ("SPSS Classification Tree"), Karten ("SPSS Maps") und interaktive Diagramme (Menü "Grafiken", Untermenü "Interaktiv") sind folgende EPS-Optionen verfügbar:

Bildvorschau. Mit dieser Option können Sie ein Vorschaubild innerhalb des EPS-Bildes speichern. Ein Vorschaubild wird vor allem dann verwendet, wenn eine EPS-Datei in einem anderen Dokument plaziert wird. Viele Anwendungen können keine EPS-Bilder, sondern lediglich die mit dem Bild gespeicherte Vorschau anzeigen. Das Vorschaubild kann im WMF-Format (kleiner und besser skalierbar) oder als TIFF-Datei (portabler und von anderen Plattformen unterstützt) vorliegen. Überprüfen Sie, welche Vorschauformate von der Anwendung unterstützt werden, in der die EPS-Grafik eingefügt werden soll.

Schriftarten. Steuert die Behandlung von TrueType-Schriftarten in EPS-Bildern.

- **Als nativen TrueType einbetten.** Bettet die meisten Schriftartdaten in die EPS-Datei ein. Die daraus folgende PostScript-Schriftart wird "Type 42" genannt. *Anmerkung*: Nicht alle PostScript-Drucker verfügen über Treiber der Stufe 3 zum Lesen von Schriftarten dieses Typs.
- **In PostScript-Schriftarten umwandeln.** Konvertiert TrueType-Schriftarten anhand der jeweiligen Schriftartpalette in PostScript-Schriftarten (Typ 1). Times New Roman wird beispielsweise zu Times, und Arial wird in Helvetica umgewandelt. *Anmerkung*: Dieses Format ist nicht empfehlenswert für interaktive Grafiken, in denen die Schriftart SPSS Marker verwendet wird (z. B. in Streudiagrammen), da es keine sinnvollen PostScript-Entsprechungen für die TrueType-Marker von SPSS gibt.
- **Schriftarten durch Kurven ersetzen.** Wandelt TrueType-Schriftarten in PostScript-Kurvendaten um. Der Text selbst kann dann nicht mehr bearbeitet werden. Dabei kommt es ebenfalls zu einem Qualitätsverlust. Diese Option ist jedoch sinnvoll, wenn Sie über einen PostScript-Drucker verfügen, der keine Schriftarten vom Typ 42 unterstützt, spezielle TrueType-Symbole jedoch beibehalten werden sollen, beispielsweise die in interaktiven Streudiagrammen verwendeten Marker.

Andere Diagramme

Für alle anderen Diagramme sind die folgenden EPS-Optionen verfügbar:

TIFF-Vorschau einschließen. Speichert eine Vorschau mit dem EPS-Bild im TIFF-Format zur Anzeige in Anwendungen, bei denen keine EPS-Bilder auf dem Bildschirm dargestellt werden können.

Schriftarten. Steuert die Behandlung von Schriftarten in EPS-Bildern.

- **Schriftarten durch Kurven ersetzen.** Wandelt Schriftarten in PostScript-Kurvendaten um. Der Text selbst kann dann nicht mehr bearbeitet werden. Diese Option ist sinnvoll, wenn die im Diagramm verwendeten Schriftarten auf dem Ausgabegerät nicht zur Verfügung stehen.
- **Schriftartreferenzen verwenden.** Wenn die im Diagramm verwendeten Schriftarten auf dem Ausgabegerät zur Verfügung stehen, werden sie verwendet. Anderenfalls verwendet das Ausgabegerät andere Schriftarten.

Exportoptionen für WMF-Diagramme

Aldus-Variable. Mit dieser Option verfügen Sie über eine gewisse Geräteunabhängigkeit (dieselbe physische Größe beim Öffnen mit 96 und 120 dpi). Dieses Format wird jedoch nicht von allen Anwendungen unterstützt.

Standard Windows. Wird von den meisten Anwendungen unterstützt, die Windows-Metadateien anzeigen können.

So legen Sie Exportoptionen für Diagramme fest

- Wechseln Sie in das Viewer-Fenster. (Klicken Sie auf eine beliebige Stelle im Fenster.)
- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Datei
  Exportieren...
- Klicken Sie bei Wahl der Option "Ausgabedokument" auf Optionen, und wählen Sie das Exportformat. Klicken Sie dann auf Optionen für Diagramm.
- Wenn Sie die Option "Nur Diagramme" ausgewählt haben, wählen Sie das Exportformat, und klicken Sie dann auf Optionen.

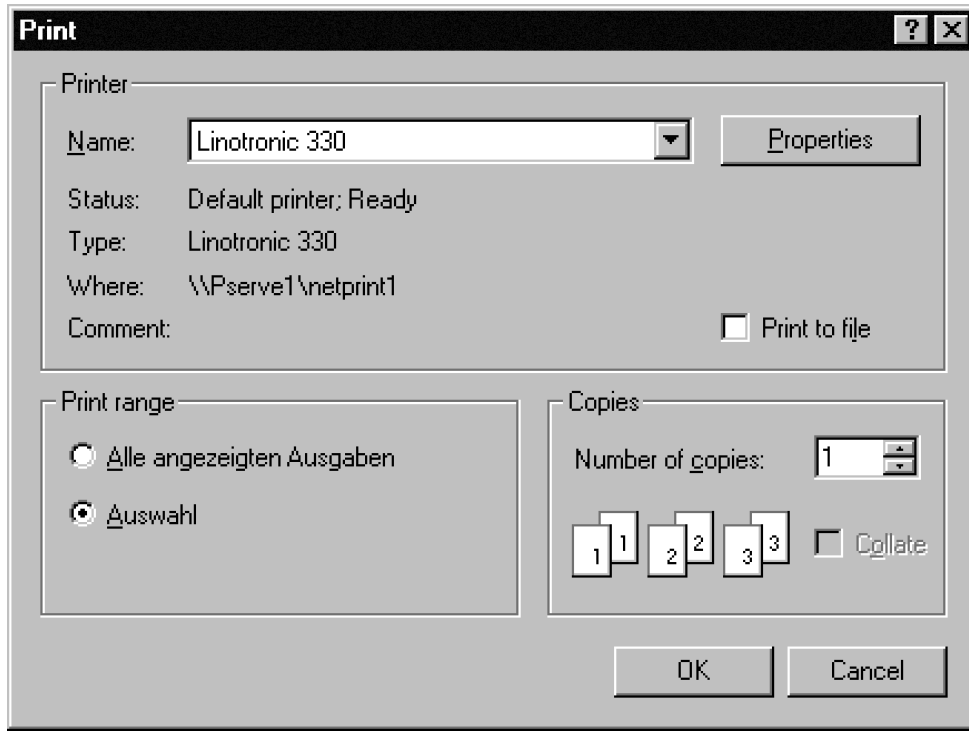
Ausdrucken von Viewer-Dokumenten

Es stehen zwei Optionen zum Drucken des Inhalts des Viewer-Fensters zur Verfügung:

Alle angezeigten Ausgaben. Hiermit werden nur die gegenwärtig im Inhaltsfenster angezeigten Objekte gedruckt. Ausgeblendete Objekte werden nicht gedruckt. (Ausgeblendete Objekte sind Objekte, die im Gliederungsfenster mit einem geschlossenen Buch dargestellt werden oder in reduzierten Gliederungsschichten verborgen sind.)

Auswahl. Hiermit werden nur die gegenwärtig im Gliederungs- und/oder Inhaltsfenster ausgewählten Objekte gedruckt.

Abbildung 10-9
Dialogfeld "Drucken" im Viewer



So drucken Sie Ausgaben und Diagramme

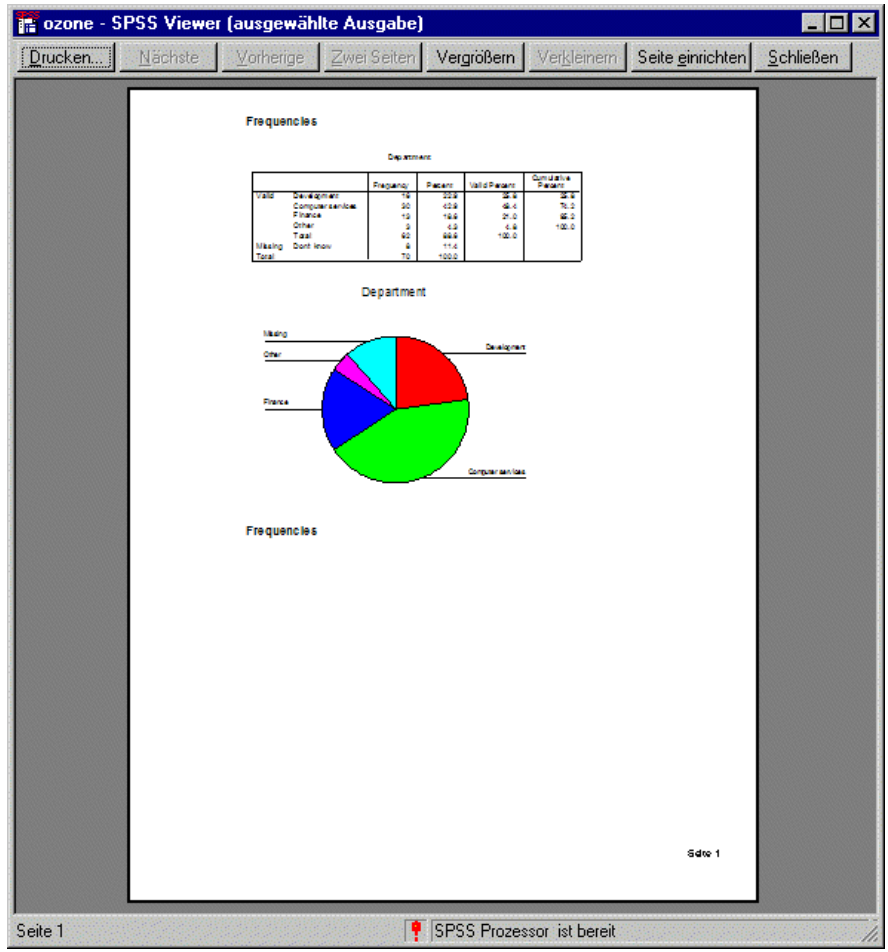
- Wechseln Sie in das Viewer-Fenster. (Klicken Sie auf eine beliebige Stelle im Fenster.)
- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Datei
  Drucken...
- Wählen Sie die gewünschten Druckereinstellungen.
- Klicken Sie zum Drucken auf OK.

Seitenansicht

Mit “Seitenansicht” erhalten Sie für jede Seite eine Vorschau des Ausdrucks von Viewer-Dokumenten. Bevor Sie Viewer-Dokumente drucken, sollten Sie diese in der Seitenansicht überprüfen. In der Seitenansicht werden Objekte angezeigt, die möglicherweise nicht sichtbar sind, wenn Sie nur das Inhaltsfenster des Viewers betrachten, beispielsweise:

- Seitenumbrüche,
- verborgene Schichten von Pivot-Tabellen,
- Umbrüche in breiten Tabellen,
- die vollständige Ausgabe aus großen Tabellen,
- auf jeder Seite zu druckende Kopf- und Fußzeilen.

Abbildung 10-10
Seitenansicht



Wenn im Viewer gegenwärtig eine Ausgabe ausgewählt ist, wird in der Seitenansicht nur diese Ausgabe angezeigt. Wenn Sie eine Vorschau für alle Ausgaben sehen möchten, darf im Viewer kein Objekt ausgewählt sein.

So lassen Sie die Seitenansicht anzeigen

- ▶ Wechseln Sie in das Viewer-Fenster. (Klicken Sie auf eine beliebige Stelle im Fenster.)

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

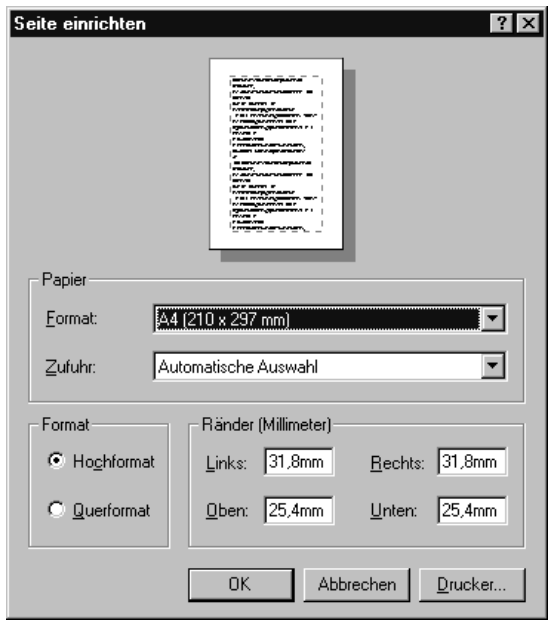
Datei
Seitenansicht

Seite einrichten

Mit dem Befehl “Seite einrichten” können Sie folgendes festlegen:

- Papierformat und -ausrichtung,
- Seitenränder,
- Kopf- und Fußzeilen,
- Seitennumerierung,
- Größe von ausgedruckten Diagrammen.

Abbildung 10-11
Dialogfeld “Seite einrichten”.



Die Einstellungen für die Seiteneinrichtung werden zusammen mit dem Viewer-Dokument gespeichert. Der Befehl “Seite einrichten” wirkt sich nur auf die Einstellungen zum Drucken von Viewer-Dokumenten aus. Diese Einstellungen

haben keine Auswirkung auf das Drucken von Daten aus dem Daten-Editor oder von Syntax aus einem Syntax-Fenster.

So ändern Sie die Seiteneinrichtung

- ▶ Wechseln Sie in das Viewer-Fenster. (Klicken Sie auf eine beliebige Stelle im Fenster.)
- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

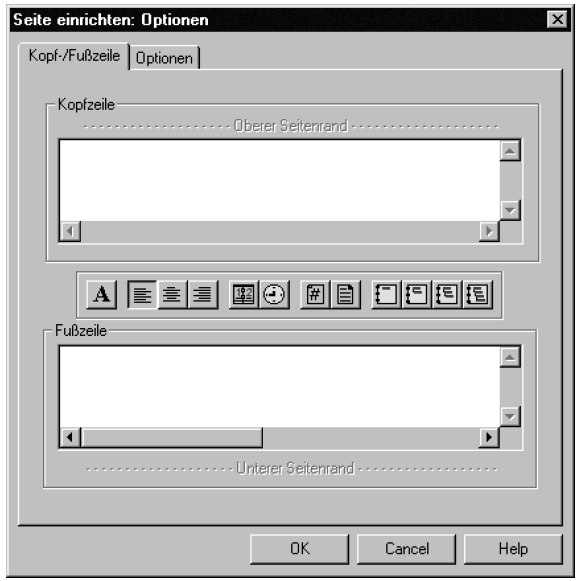
  Datei
  Seite einrichten...
- ▶ Ändern Sie die Einstellungen, und klicken Sie auf OK.

Seite einrichten: Optionen, Kopf-/Fußzeile

Kopf- und Fußzeilen sind die Informationen, die am oberen und unteren Rand jeder Seite ausgedruckt werden. Sie können beliebigen Text als Kopf- und Fußzeile eingeben. Außerdem können Sie die Symbolleiste in der Mitte des Dialogfelds verwenden, wenn Sie folgendes einfügen möchten:

- Datum und Uhrzeit,
- Seitennummern,
- Dateiname aus dem Viewer,
- Beschriftungen der Gliederungsüberschriften,
- Titel und Untertitel.

Abbildung 10-12
Dialogfeld "Seite einrichten: Optionen", Registerkarte "Kopf-/Fußzeile"



Mit den Beschriftungen der Gliederungsüberschriften werden die Gliederungsüberschriften der ersten, zweiten, dritten und/oder vierten Stufe für das erste Objekt auf jeder Seite angezeigt.

Mit den Symbolen für die Seitentitel und die Untertitel werden die aktuellen Titel und Untertitel der Seite gedruckt. Seitentitel und Untertitel werden mit der Option "Neuer Seitentitel" im Menü "Einfügen" des Viewersoder mit den Befehlen `TITLE` und `SUBTITLE` in der Befehlssyntax erstellt. Wenn Sie keine Seitentitel oder -untertitel angegeben haben, wird diese Einstellung ignoriert.

Anmerkung: Die Merkmale der Schriftart für neue Seitentitel und -untertitel werden auf der Registerkarte "Viewer" des Dialogfelds "Optionen" festgelegt (Menü "Bearbeiten", Befehl "Optionen"). Bei bestehenden Seitentiteln und –untertiteln können die Merkmale der Schriftart durch Bearbeiten der Titel direkt im Viewer festgelegt werden.

Um eine Vorschau des Erscheinungsbildes von Kopf- und Fußzeilen auf der gedruckten Seite zu erhalten, wählen Sie im Menü "Datei" die Option Seitenansicht aus.

Seite einrichten: Optionen, Optionen

In diesem Dialogfeld werden die Größe der gedruckten Diagramme, der Abstand zwischen gedruckten Ausgabeobjekten und die Seitennumerierung eingestellt.

Größe des gedruckten Diagramms. Hier wird die Größe des gedruckten Diagramms im Verhältnis zur definierten Seitengröße festgelegt. Die Größe des gedruckten Diagramms hat keine Auswirkungen auf das Seitenverhältnis (Verhältnis Breite zu Höhe) des Diagramms. Die Gesamtgröße eines gedruckten Diagramms wird von dessen Höhe und Breite bestimmt. Wenn die äußeren Grenzen eines Diagramms den linken und rechten Rand der Seite berühren, kann das Diagramm in der Höhe nicht weiter vergrößert werden.

Abstand zwischen Objekten. Hier wird der Abstand zwischen gedruckten Objekten festgelegt. Jede Pivot-Tabelle, jedes Diagramm und jedes Textobjekt ist ein separates Objekt. Diese Einstellung wirkt sich nicht auf die Anzeige von Objekten im Viewer aus.

Seitennumerierung beginnen mit. Die Seiten werden fortlaufend ab der angegebenen Nummer numeriert.

Abbildung 10-13
Dialogfeld "Seite einrichten: Optionen", Registerkarte "Optionen"

Speichern der Ausgabe

Der Inhalt des Viewers kann in einem Viewer-Dokument gespeichert werden. Das gespeicherte Dokument enthält beide Bereiche des Viewer-Fensters (Gliederung und Inhalt).

So speichern Sie ein Viewer-Dokument

- Wählen Sie die folgenden Befehle aus den Menüs des Viewer-Fensters aus:

  Datei
  Speichern

- Geben Sie einen Namen für das Dokument ein, und klicken Sie anschließend auf Speichern.

  Verwenden Sie zum Speichern von Ergebnissen in externen Formaten (zum Beispiel HTML-Format oder Text-Format) den Befehl “Exportieren” im Menü “Datei”. (Diese Funktion ist im selbstständig ausführbaren SmartViewer nicht verfügbar.)

Option “Mit Paßwort speichern”

Mit dem Befehl “Mit Paßwort speichern” können Sie Viewer-Dateien mit einem Paßwort-Schutz versehen.

Paßwort. Beim Passwort muss die Groß- und Kleinschreibung beachtet werden. Es kann bis zu 16 Zeichen lang sein. Wenn Sie ein Passwort zuweisen, kann die Datei nicht ohne Eingabe des Passworts betrachtet werden.

OEM-Code. Lassen Sie dieses Feld frei, wenn Sie keine vertragliche Vereinbarung mit SPSS über die Weitergabe des SmartViewers getroffen haben. Der OEM-Lizenzcode wird mit dem Vertrag zur Verfügung gestellt.

So speichern Sie Viewer-Dateien mit einem Passwort

- Wählen Sie die folgenden Befehle aus den Menüs des Viewer-Fensters aus:

  Datei
  Mit Passwort speichern...

- Geben Sie das Passwort ein.

- Geben Sie das Passwort zur Bestätigung erneut ein, und klicken Sie anschließend auf OK.
- Geben Sie im Dialogfeld “Speichern unter” einen Dateinamen ein.
- Klicken Sie auf Speichern.

Anmerkung: Lassen Sie das Feld “OEM-Code” frei, wenn Sie keine vertragliche Vereinbarung mit SPSS über die Weitergabe des SmartViewers getroffen haben.

Kapitel
11

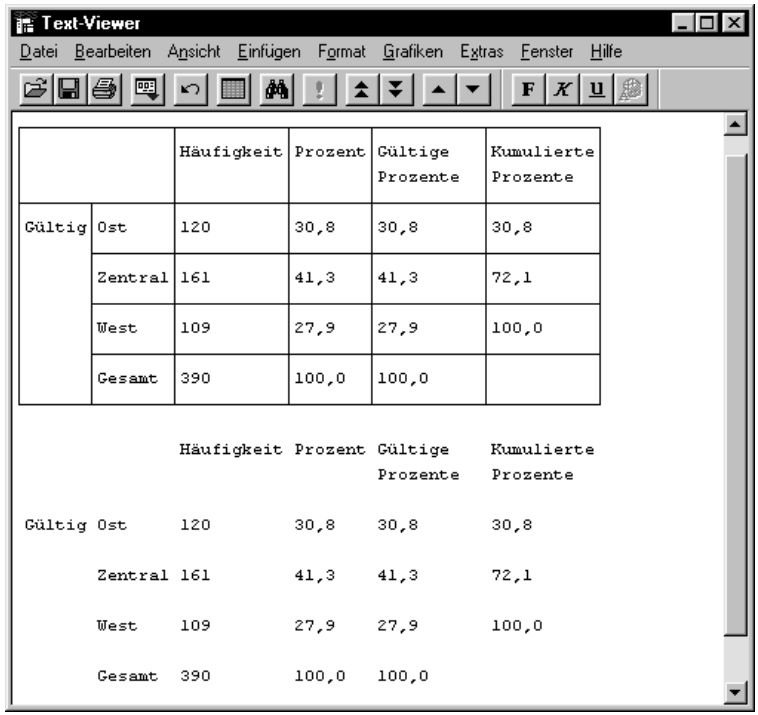
Text-Viewer

Der Text-Viewer stellt Ergebnisse in “Textform” dar. Dies beinhaltet:

- Ausgabe von einfachem Text (anstelle von Pivot-Tabellen),
- Diagramme im Metadatei-Bildformat (anstelle von Diagrammobjekten).

Im Text-Viewer können Sie Textausgaben bearbeiten und Diagrammgrößen ändern sowie Textausgaben und Diagramme in andere Anwendungen einfügen. Diagramme können jedoch nicht bearbeitet werden. Außerdem sind die interaktiven Funktionen für Pivot-Tabellen und Diagramme nicht verfügbar.

Abbildung 11-1
Text-Viewer-Fenster



		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Ost	120	30,8	30,8	30,8
	Zentral	161	41,3	41,3	72,1
	West	109	27,9	27,9	100,0
	Gesamt	390	100,0	100,0	

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Ost	120	30,8	30,8	30,8
	Zentral	161	41,3	41,3	72,1
	West	109	27,9	27,9	100,0
	Gesamt	390	100,0	100,0	

So erstellen Sie eine Textausgabe:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Datei
  Neu
  Textausgabe

- Um die Textausgabe als Standard-Ausgabetyp festzulegen, wählen Sie folgende Befehle aus den Menüs aus:

  Bearbeiten
  Optionen

- Klicken Sie auf die Registerkarte Allgemein.

- Wählen Sie unter “Viewer-Typ beim Start” die Option Text.

Anmerkung: Neue Ausgabe wird immer im Hauptfenster des Text-Viewers angezeigt. Wenn Sie sowohl ein Viewer- als auch ein Text-Viewer-Fenster geöffnet haben, ist das zuletzt geöffnete das **Hauptfenster**. Sie können diese Eigenschaft auch mit der Schaltfläche "Hauptfenster" (das Symbol mit dem Ausrufungszeichen) auf der Symbolleiste zuweisen.

Steuern des Textausgabeformats

Ausgaben, die im Viewer als Pivot-Tabellen dargestellt würden, werden im Text-Viewer in Textausgaben umgewandelt. Die Standardeinstellungen für die Ausgabe umgewandelter Pivot-Tabellen umfassen:

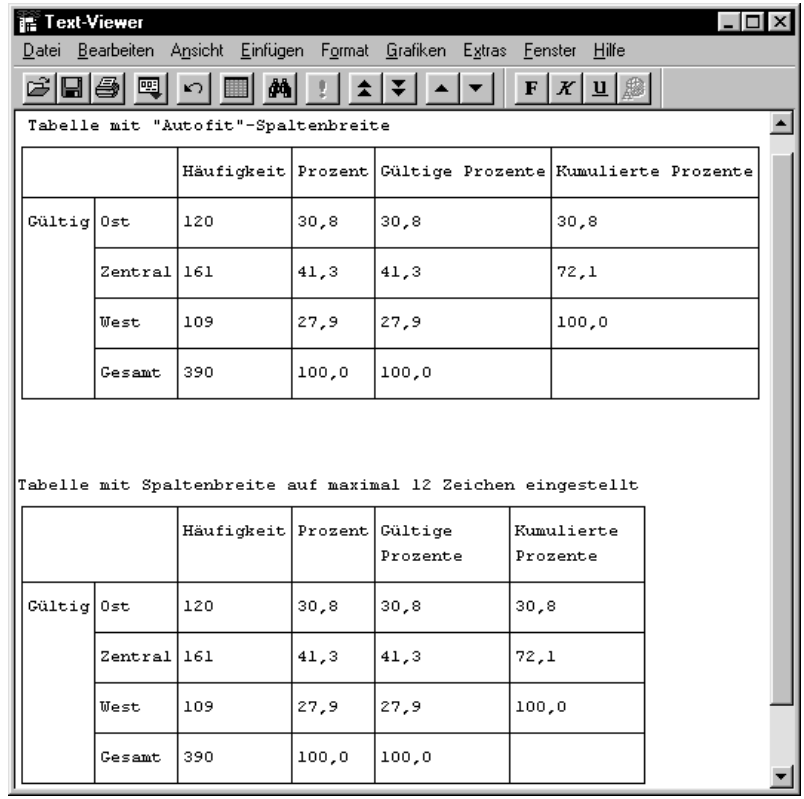
- Die Spaltenbreite richtet sich nach der Breite der Spaltenbeschriftung. Außerdem werden Beschriftungen nicht in mehrere Zeilen umgebrochen.
- Die Ausrichtung wird mit Leerzeichen gesteuert (nicht mit Tabulatoren).
- Als Trennzeichen für Zeilen und Spalten werden Rahmenlinien der Schriftart SPSS Marker Set verwendet.
- Wenn Rahmenlinien deaktiviert sind, werden als Trennzeichen vertikale Linien (|) für Spalten und Bindestriche (–) für Zeilen eingesetzt.

Sie können das Format der neuen Testausgabe auf der Registerkarte "Text-Viewer" festlegen. Wählen Sie dazu im Menü "Bearbeiten" den Befehl "Optionen" und im dann angezeigten Dialogfeld "Optionen" die Registerkarte "Text-Viewer".

Abbildung 11-2
Optionen: Text-Viewer

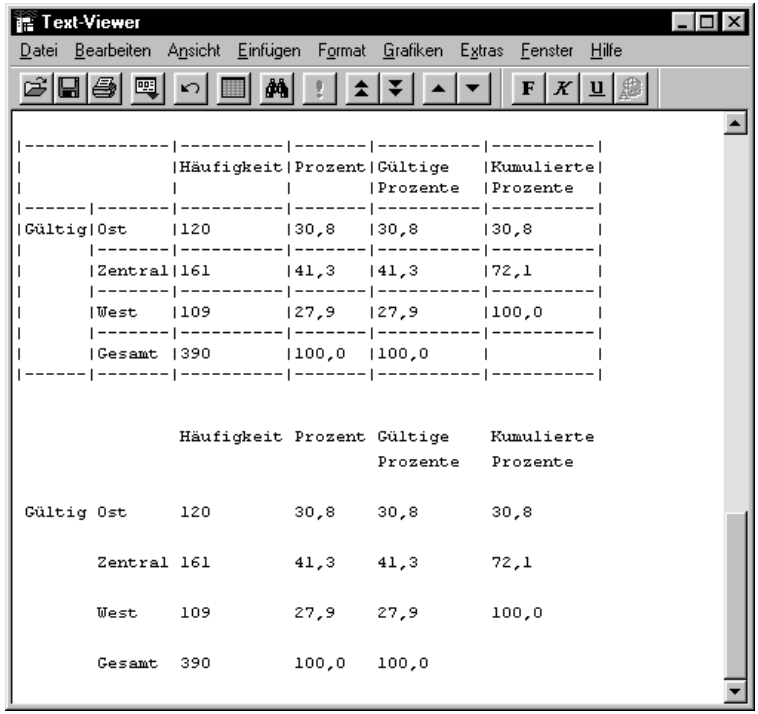
Spaltenbreite. Wenn Tabellen lange Spaltenbeschriftungen beinhalten, wählen Sie für die Spaltenbreite Maximale Zeichenanzahl aus, um die Breite solcher Tabellen zu verringern. Wenn Beschriftungen die festgelegte Breite überschreiten, werden sie umbrochen.

Abbildung 11-3
Textausgabe vor und nach dem Einstellen der maximalen Spaltenbreite


Text-Viewer
Datei Bearbeiten Ansicht Einfügen Format Grafiken Extras Fenster Hilfe
Tabelle mit "Autofit"-Spaltenbreite
Häufigkeit Prozent Gültige Prozente Kumulierte Prozente
Gültig Ost 120 30,8 30,8 30,8
Zentral 161 41,3 41,3 72,1
West 109 27,9 27,9 100,0
Gesamt 390 100,0 100,0
Tabelle mit Spaltenbreite auf maximal 12 Zeichen eingestellt
Häufigkeit Prozent Gültige Prozente Kumulierte Prozente
Gültig Ost 120 30,8 30,8 30,8
Zentral 161 41,3 41,3 72,1
West 109 27,9 27,9 100,0
Gesamt 390 100,0 100,0

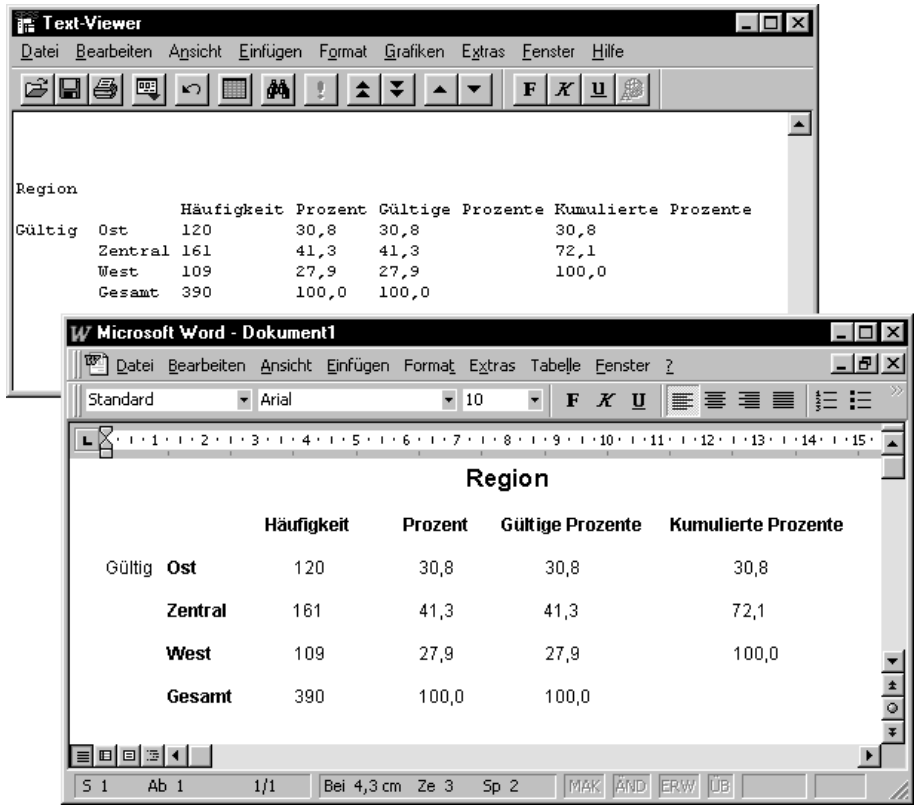
Zeilen- und Spaltentrennzeichen. Anstelle von Rahmenlinien für Zeilen und Spalten können Sie in den Einstellungen für Zellentrennzeichen die Zeilen- und Spaltentrennzeichen für die Ausgabe von neuem Text festlegen. Sie können verschiedene Zellentrennzeichen festlegen oder Leerzeichen eingeben, wenn Sie zum Markieren von Zeilen und Spalten bestimmte Zeichen nicht verwenden möchten. Zur Angabe von Zellentrennzeichen müssen Sie Boxzeichen anzeigen deaktivieren.

Abbildung 11-4
Textausgabe vor und nach dem Einstellen von Zellentrennzeichen


Text-Viewer
Datei Bearbeiten Ansicht Einfügen Format Grafiken Extras Fenster Hilfe
|--------------|----------|-------|----------|----------|
| |Häufigkeit|Prozent|Gültige |Kumulierte|
			Prozente	Prozente
Gültig	Ost	120	30,8	30,8	30,8
	-------	----------	-------	----------	----------
	Zentral	161	41,3	41,3	72,1
	-------	----------	-------	----------	----------
	West	109	27,9	27,9	100,0
	-------	----------	-------	----------	----------
	Gesamt	390	100,0	100,0	
Häufigkeit Prozent Gültige Kumulierte
Prozente Prozente
Gültig Ost 120 30,8 30,8 30,8
Zentral 161 41,3 41,3 72,1
West 109 27,9 27,9 100,0
Gesamt 390 100,0 100,0

Leerzeichen-getrennte Spalten und Tabulator-getrennte Spalten. Der Text-Viewer zeigt Leerzeichen-getrennte Ausgaben im nicht proportionalen Zeichensatz an. Wenn Sie eine Textausgabe in eine andere Anwendung einfügen, können Sie nur einen nicht proportionalen Zeichensatz verwenden, damit durch Leerzeichen getrennte Spalten korrekt ausgerichtet werden. Verwenden Sie hingegen Tabulatoren als Spaltentrennzeichen, können Sie in der anderen Anwendung mit einem beliebigen Zeichensatz arbeiten und Tabulatoren für die korrekte Ausrichtung der Ausgabe verwenden. Durch Tabulator getrennte Ausgaben werden allerdings im Text-Viewer nicht korrekt ausgerichtet.

Abbildung 11-5
Durch Tabulator getrennte Ausgaben im Text-Viewer, formatiert in einer Textverarbeitung



So legen Sie Optionen für den Text-Viewer fest

- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Bearbeiten
  Optionen...
- Klicken Sie auf die Registerkarte Text-Viewer.
- Wählen Sie die gewünschten Einstellungen aus.
- Klicken Sie auf OK oder Zuweisen.

Die Optionen für die Anzeige der Ausgaben im Text-Viewer wirken sich nur auf die Ausgaben aus, die nach dem Ändern der Einstellungen berechnet werden. Änderungen der Einstellungen wirken sich nicht auf die im Text-Viewer bereits angezeigte Ausgabe aus.

Eigenschaften der Schriftart

Sie können die Eigenschaften der Schriftart (z. B. Art, Stil, Größe) von Textausgaben im Text-Viewer ändern. Wenn Sie in einem durch Leerzeichen getrenntem Text Rahmenlinien für Zeilen und Spalten einsetzen, müssen Sie für die korrekte Ausrichtung der Spalten einen nicht proportionalen Zeichensatz wie beispielsweise Courier verwenden. Änderungen an der Schriftart, wie beispielsweise an Größe und Stil (z. B. fett und kursiv), können sich auf die Spaltenausrichtung auswirken, wenn die Änderungen sich nur auf einen bestimmten Teil einer Tabelle beziehen.

Zeilen- und Spaltenrahmen. In der Standardeinstellung werden durchgezogene Zeilen- und Spaltenrahmen in der Schriftart SPSS Marker Set verwendet. Die für die Rahmen verwendeten Zeichen werden von anderen Schriftarten nicht unterstützt.

So ändern Sie die Eigenschaften der Schriftart für die Textausgabe

- Wählen Sie den Text, dem Sie eine Änderung an der Schriftart zuweisen möchten.
- Wählen Sie die folgenden Befehle aus den Text-Viewer-Menüs aus:

  Format
  Schriftart...
- Wählen Sie die Eigenschaften für die Schriftart aus, die Sie auf den markierten Text anwenden möchten.

So drucken Sie eine Textausgabe

- Wählen Sie die folgenden Befehle aus den Text-Viewer-Menüs aus:

  Datei
  Drucken...

So können Sie einen bestimmten Abschnitt der Textausgabe drucken

- Wählen Sie die Ausgabe, die gedruckt werden soll.

- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Datei
  Drucken...

- Aktivieren Sie das Optionsfeld Auswahl.

Seitenansicht

Mit “Seitenansicht” erhalten Sie eine seitenweise Vorschau auf den Ausdruck von Text-Dokumenten. Vor dem Drucken eines Text-Dokuments sollten Sie die Seitenansicht überprüfen, weil in der Seitenansicht Objekte angezeigt werden, die möglicherweise nicht auf die Seite passen, zum Beispiel:

- Lange Tabellen,
- Breite Tabellen aus umgewandelten Ausgaben von Pivot-Tabellen ohne Anpassung der Spaltenbreite,
- Textausgaben, die mit der Option “Breit” für die Seitenbreite (Dialogfeld “Optionen”, Registerkarte “Text-Viewer”) erstellt wurden, wobei der Drucker auf “Hochformat” eingestellt ist.

Eine Ausgabe, die zu breit für die Seite ist, wird abgeschnitten und nicht auf einer anderen Seite gedruckt. Sie verfügen über mehrere Möglichkeiten, das Abschneiden von zu breiter Ausgabe zu verhindern:

- Verwenden Sie eine kleinere Schriftarten. Wählen Sie dazu im Menü “Format” den Befehl “Schriftarten”.
- Wählen Sie als Seitenausrichtung Querformat (Menü “Datei”, Befehl “Seite einrichten”).
- Geben Sie für die neue Ausgabe eine schmale maximale Spaltenbreite an. Wählen Sie dazu im Menü “Bearbeiten” den Befehl “Optionen” und im angezeigten Dialogfeld die Registerkarte “Text-Viewer”.

Verwenden Sie für lange Tabellen Seitenumbrüche, um genau festzulegen, wo die Tabelle zwischen den Seiten umbrochen wird. Wählen Sie dazu im Menü “Einfügen” den Befehl “Seitenumbruch”.

So lassen Sie sich die Textausgabe in der Seitenansicht anzeigen

- Wählen Sie die folgenden Befehle aus den Text-Viewer-Menüs aus:

  Datei
  Seitenansicht

So speichern Sie Text-Viewer-Ausgaben

- Wählen Sie die folgenden Befehle aus den Text-Viewer-Menüs aus:

  Datei
  Speichern

  Text-Viewer-Ausgaben werden im RTF-Format (Rich Text Format) gespeichert.

So speichern Sie die Textausgabe im Textformat

- Wählen Sie die folgenden Befehle aus den Text-Viewer-Menüs aus:

  Datei
  Exportieren...

  Sie können den ganzen Text oder ausgewählte Textabschnitte exportieren. Es werden nur Textausgaben (Ausgaben umgewandelter Pivot-Tabellen und Textausgaben) in exportierten Dateien gespeichert, Diagramme werden nicht eingeschlossen.

Kapitel
12

Pivot-Tabellen

Viele Ergebnisse im Viewer werden in Tabellen dargestellt, die Sie interaktiv pivotieren können. Das heißt, Sie können die Zeilen, Spalten und Schichten neu anordnen.

Bearbeiten von Pivot-Tabellen

Für die Bearbeitung von Pivot-Tabellen stehen folgende Optionen zur Verfügung:

- Transponieren von Zeilen und Spalten,
- Verschieben von Zeilen und Spalten,
- Erstellen von mehrdimensionalen Schichten,
- Anlegen und Aufheben von Gruppierungen für Zeilen und Spalten,
- Ein- und Ausblenden von Zellen,
- Drehen von Zeilen- und Spaltenbeschriftungen,
- Anzeigen von Definitionen für Terme.

So bearbeiten Sie eine Pivot-Tabelle

▶ Doppelklicken Sie auf die Tabelle.

Dadurch wird der Pivot-Tabellen-Editor gestartet.

So bearbeiten Sie mehrere Pivot-Tabellen gleichzeitig

▶ Klicken Sie mit der rechten Maustaste auf die Pivot-Tabelle.

- Wählen Sie folgende Menübefehle aus dem Kontextmenü aus:

  SPSS-Pivot-Tabellenobjekt
  Öffnen

- Wiederholen Sie diesen Vorgang für jede Pivot-Tabelle, die Sie bearbeiten möchten.

  Anschließend können Sie die Pivot-Tabellen jeweils in einem separaten Fenster bearbeiten.

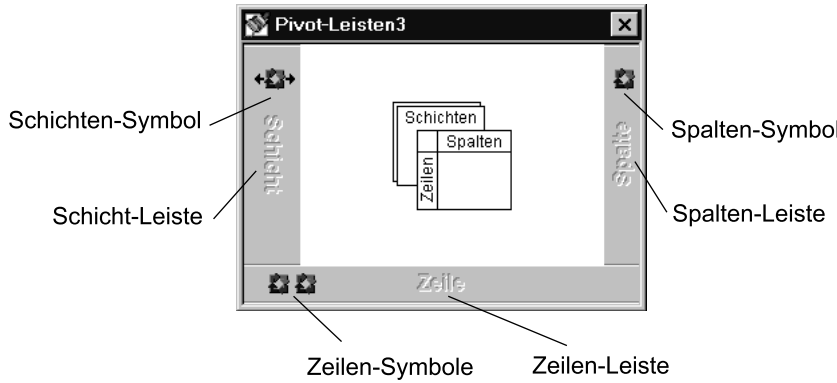
So pivotieren Sie eine Tabelle mithilfe von Symbolen

- Aktivieren Sie die Pivot-Tabelle.
- Wählen Sie die folgenden Befehle aus den Menüs für Pivot-Tabellen aus:

  Pivot
  Pivot-Leisten

- Wenn Sie mit dem Mauszeiger auf eines der Symbole zeigen, wird in einer QuickInfo angezeigt, welche Dimension das Symbol darstellt.
- Ziehen Sie ein Symbol von einer Leiste auf eine andere Leiste.

Abbildung 12-1
Pivot-Leisten



Dadurch wird die Anordnung der Tabelle geändert. Angenommen, das Symbol stellt eine Variable mit den Kategorien Ja und Nein dar. Wenn Sie das Symbol von der Zeilen-Leiste auf die Spalten-Leiste ziehen, werden Ja und Nein zu Spaltenbeschriftungen. Vor dem Verschieben waren Ja und Nein Zeilenbeschriftungen.

So bestimmen Sie die Dimensionen in Pivot-Tabellen

- Aktivieren Sie die Pivot-Tabelle.
- Wenn die Pivot-Leisten noch nicht eingeblendet sind, wählen Sie folgende Befehle aus den Menüs aus:

  Pivot
  Pivot-Leisten
- Klicken Sie auf ein Symbol und halten Sie die Maustaste gedrückt.

  Dadurch werden die zugehörigen Dimensionsbeschriftungen in der Pivot-Tabelle hervorgehoben.

So transponieren Sie Zeilen und Spalten

- Wählen Sie die folgenden Befehle aus den Menüs für Pivot-Tabellen aus:

  Pivot
  Zeilen und Spalten vertauschen

  Damit erzielen Sie dasselbe Ergebnis, als wenn Sie alle Zeilen-Symbole in die Spalten-Leiste und alle Spalten-Symbole in die Zeilen-Leiste ziehen würden.

So ändern Sie die Anzeigereihenfolge

Die Reihenfolge von Pivot-Symbolen in einer Dimensions-Leiste spiegelt die Anzeigereihenfolge der Elemente in der Pivot-Tabelle wider. So ändern Sie die Anzeigereihenfolge von Elementen in einer Dimension:

- Aktivieren Sie die Pivot-Tabelle.
- Wenn die Pivot-Leisten noch nicht eingeblendet sind, wählen Sie folgende Befehle aus den Menüs aus:

  Pivot
  Pivot-Leisten
- Ziehen Sie die Symbole in den einzelnen Leisten in die gewünschte Reihenfolge (von links nach rechts oder von oben nach unten).

So verschieben Sie Zeilen und Spalten in einer Pivot-Tabelle

- Aktivieren Sie die Pivot-Tabelle.
- Klicken Sie auf die Beschriftung der zu verschiebenden Zeile oder Spalte.
- Ziehen Sie die Beschriftung an die neue Position.
- Wählen Sie im Kontextmenü den Befehl Einfügen vor oder Vertauschen aus.

Anmerkung: Stellen Sie sicher, daß die Option "Zum Kopieren ziehen" im Menü "Bearbeiten" *nicht* aktiviert ist (also nicht mit einem Häkchen gekennzeichnet ist). Wenn die Option "Zum Kopieren ziehen" aktiviert ist, deaktivieren Sie diese.

So gruppieren Sie Zeilen oder Spalten und fügen Gruppenbeschriftungen ein

- Aktivieren Sie die Pivot-Tabelle.
- Wählen Sie die Beschriftungen für die Zeilen oder Spalten aus, die Sie gruppieren möchten. (Klicken und ziehen Sie mit der Maus, oder halten Sie beim Klicken die Umschalt-Taste gedrückt, um mehrere Beschriftungen auszuwählen).
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Gruppieren

Es wird automatisch eine Gruppenbeschriftung eingefügt. Doppelklicken Sie auf die Gruppenbeschriftung, um den Text zu bearbeiten.

Abbildung 12-2
Zeilen- und Spaltengruppen und Beschriftungen

		Column Group Label		
		Female	Male	Total
Row Group Label	Clerical	206	157	363
	Custodial		27	27
	Manager	10	74	84

Anmerkung: Wenn Sie Zeilen oder Spalten zu einer vorhandenen Gruppe hinzufügen möchten, müssen Sie zuerst die entsprechende Gruppierung aufheben und dann eine neue Gruppe anlegen, welche die zusätzlichen Objekte enthält.

So heben Sie die Gruppierung von Zeilen oder Spalten auf und entfernen Gruppenbeschriftungen

- ▶ Aktivieren Sie die Pivot-Tabelle.
- ▶ Klicken Sie auf eine beliebige Stelle in der Gruppenbeschriftung der Zeilen oder Spalten, deren Gruppierung Sie aufheben möchten.
- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Gruppierung aufheben

  Beim Aufheben der Gruppierung wird automatisch die Gruppenbeschriftung gelöscht.

So drehen Sie Beschriftungen von Pivot-Tabellen

- ▶ Aktivieren Sie die Pivot-Tabelle.
- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

  Format
  Innere Spaltenbeschriftungen drehen

  oder

  Äußere Zeilenbeschriftungen drehen

Abbildung 12-3
Gedrehte Spaltenbeschriftungen

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Büro	363	76,6	76,6	76,6
Bewachung	27	5,7	5,7	82,3
Management				
Gesamt	4			

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Büro	363	76,6	76,6	76,6
Bewachung	27	5,7	5,7	82,3
Management	84	17,7	17,7	100,0
Gesamt	474	100,0	100,0	

Nur die innersten Spaltenbeschriftungen und die äußersten Zeilenbeschriftungen können gedreht werden.

So setzen Sie Pivot-Tabellen auf die Standardwerte zurück

Nach dem Durchführen von einem oder mehreren Pivotiervorgängen können Sie die ursprüngliche Anordnung der Pivot-Tabelle wiederherstellen.

- Wählen Sie hierfür aus dem Menü "Pivot" den Befehl Pivots auf Standardwerte zurücksetzen.

Dabei werden nur Änderungen rückgängig gemacht, die durch das Pivotieren von Zeilen-, Spalten- und Schichtelementen zwischen den Dimensionen entstanden sind. Dieser Vorgang wirkt sich nicht auf Änderungen wie das Gruppieren, Aufheben von Gruppierungen oder Verschieben von Zeilen und Spalten aus.

So finden Sie Definitionen für Beschriftungen in Pivot-Tabellen auf

Sie können kontextsensitive Hilfe zu den Zellenbeschriftungen in Pivot-Tabellen erhalten. Wenn zum Beispiel *Mittelwert* als Beschriftung einer Zelle angezeigt wird, können Sie eine Definition für den Mittelwert ausgeben lassen.

- Klicken Sie mit der rechten Maustaste auf eine Zelle mit einer Beschriftung.
- Wählen Sie im Kontextmenü die Option Direkthilfe aus.

Sie müssen dabei mit der rechten Maustaste direkt auf die Zelle mit der Beschriftung klicken, nicht auf die Datenzellen in den Zeilen oder Spalten. Die kontextsensitive Hilfe steht jedoch nicht bei benutzerdefinierten Beschriftungen zur Verfügung, wie zum Beispiel Variablennamen oder Wertelabels.

Arbeiten mit Schichten

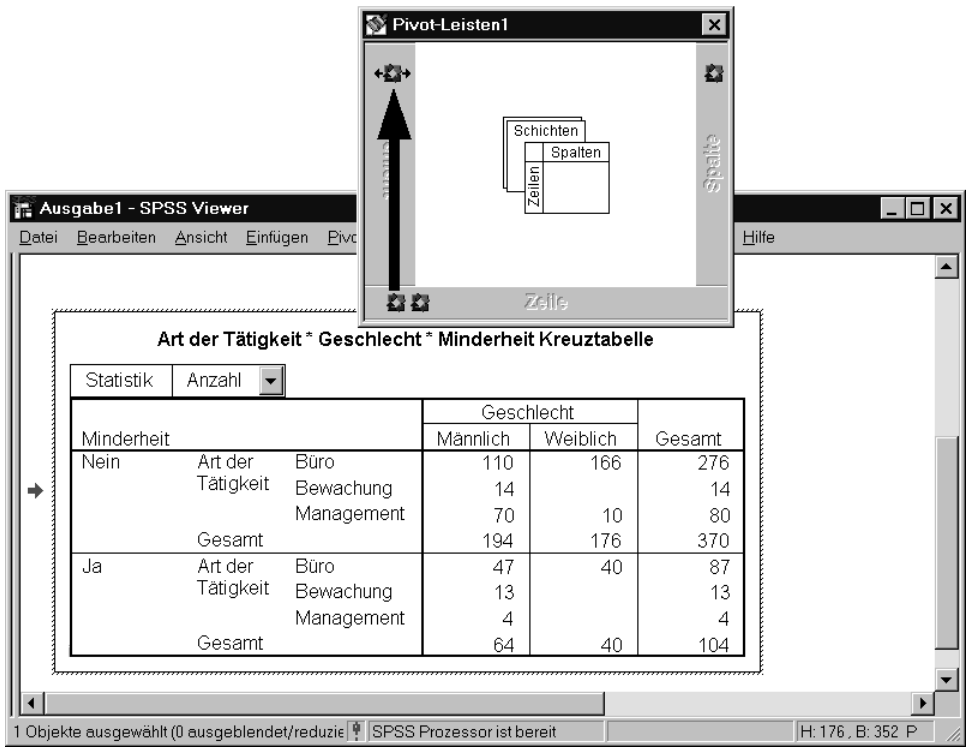
Sie können eine separate zweidimensionale Tabelle für jede Kategorie oder Kombination von Kategorien anzeigen lassen. Die Tabelle ist dabei sozusagen in Schichten gestapelt, und nur die oberste Schicht ist sichtbar.

So erstellen Sie Schichten und lassen diese anzeigen

- Aktivieren Sie die Pivot-Tabelle.

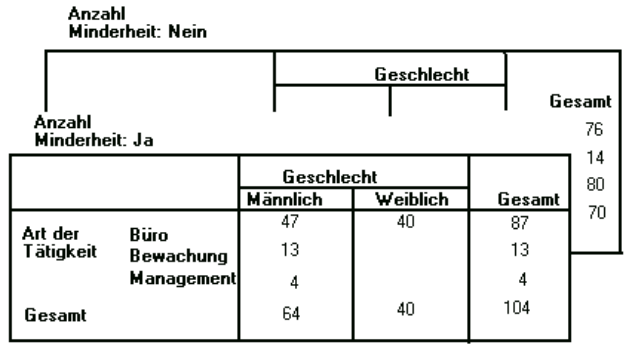
- Wählen Sie aus dem Menü “Pivot” die Option Pivot-Leisten aus, sofern diese Option nicht bereits ausgewählt ist.
- Ziehen Sie ein Symbol aus der Zeilen-Leiste oder Spalten-Leiste in die Schicht-Leiste.

Abbildung 12-4
Verschieben von Kategorien in Schichten



Jedes Schichten-Symbol weist Pfeile auf, die nach links und nach rechts zeigen. Die sichtbare Tabelle ist die Tabelle für die oberste Schicht.

Abbildung 12-5
Kategorien in separaten Schichten



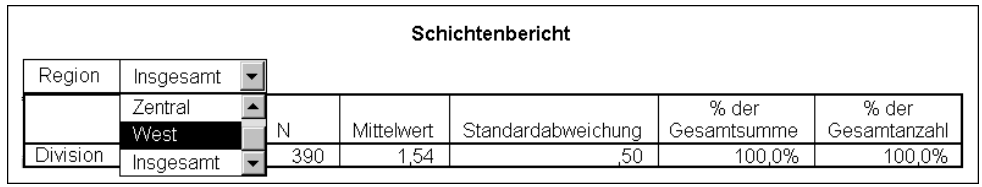
So wechseln Sie die angezeigte Schicht

- Klicken Sie auf einen der Pfeile des Schichten-Symbols.

  oder

- Wählen Sie eine Kategorie aus der Dropdown-Liste mit den Schichten aus.

Abbildung 12-6
Auswählen von Schichten aus Dropdown-Listen



Gehe zu Kategorie in Schicht

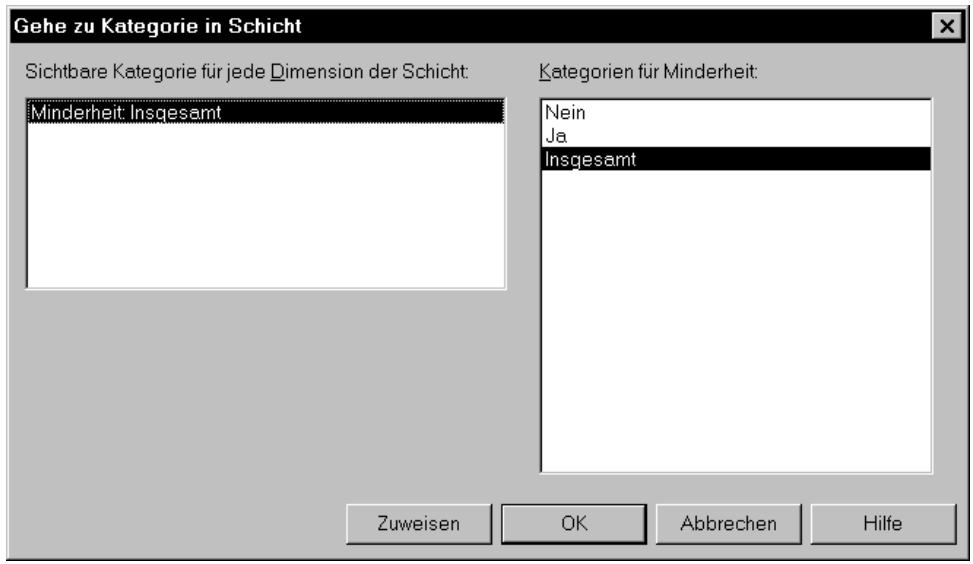
Mit “Gehe zu Kategorie in Schicht” können Sie zu einer anderen Schicht in einer Pivot-Tabelle wechseln. Dieses Dialogfeld ist besonders dann nützlich, wenn viele Schichten vorhanden sind oder eine Schicht viele Kategorien aufweist.

So wechseln Sie zu einer Tabellenschicht

- Wählen Sie die folgenden Befehle aus den Menüs für Pivot-Tabellen aus:

  Pivot
  Gehe zu Schicht...

  Abbildung 12-7
  Dialogfeld "Gehe zu Kategorie in Schicht"



- Wählen Sie aus der Liste "Sichtbare Kategorie" eine Schichtendimension aus. In der Liste "Kategorien für Schichten" werden alle Kategorien für die ausgewählte Dimension angezeigt.

- Wählen Sie aus der Liste "Kategorien" die gewünschte Kategorie aus, und klicken Sie dann auf OK. Die Schicht wird gewechselt, und das Dialogfeld wird geschlossen.

  So lassen Sie eine andere Schicht anzeigen, ohne das Dialogfeld zu schließen:

- Wählen Sie die Kategorie aus, und klicken Sie auf Zuweisen.

So verschieben Sie Schichten in Zeilen oder Spalten

Wenn die angezeigte Tabelle in Schichten gestapelt und nur die oberste Schicht sichtbar ist, können Sie alle Schichten gleichzeitig anzeigen lassen, entweder entlang der Zeilen oder entlang der Spalten. Es muß sich mindestens ein Symbol in der Schichten-Leiste befinden.

- ▶ Wählen Sie im Menü "Pivot" den Befehl Schichten in Zeilen verschieben aus.

  oder

- ▶ Wählen Sie im Menü "Pivot" den Befehl Schichten in Spalten verschieben aus.

  Sie können außerdem Schichten in Zeilen oder Spalten verschieben, indem Sie die entsprechenden Symbole zwischen den Schicht-, Zeilen- und Spalten-Leisten verschieben.

Lesezeichen

Mit Hilfe von Lesezeichen können Sie verschiedene Ansichten von Pivot-Tabellen speichern. Mit Lesezeichen werden folgende Merkmale gespeichert:

- ■ Anordnung der Elemente in den Zeilen-, Spalten- und Schichtdimensionen,
- ■ Anzeigereihenfolge der Elemente in jeder Dimension,
- ■ Gegenwärtig angezeigte Schicht für jedes Schichtelement.

So versehen Sie Pivot-Tabellen-Ansichten mit einem Lesezeichen

- ▶ Aktivieren Sie die Pivot-Tabelle.
- ▶ Pivotieren Sie die Tabelle so, dass die Ansicht angezeigt wird, die Sie mit einem Lesezeichen versehen möchten.
- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

  Pivot
  Lesezeichen…
- ▶ Geben Sie einen Namen für das Lesezeichen ein. (Bei Namen für Lesezeichen wird nicht zwischen Groß- und Kleinschreibung unterschieden.)
- ▶ Klicken Sie auf Hinzufügen.

Für jede Pivot-Tabelle gilt ein eigenes Set von Lesezeichen. Innerhalb einer Pivot-Tabelle muß jeder Name für Lesezeichen eindeutig sein. Bei verschiedenen Pivot-Tabellen können Sie Namen für Lesezeichen jedoch doppelt vergeben.

So lassen Sie eine mit einem Lesezeichen versehene Pivot-Tabellen-Ansicht anzeigen

- ▶ Aktivieren Sie die Pivot-Tabelle.
- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:
  Pivot
  Lesezeichen…
- ▶ Klicken Sie in der Liste auf den Namen des Lesezeichens.
- ▶ Klicken Sie auf Gehe zu.

So benennen Sie Lesezeichen für Pivot-Tabellen um

- ▶ Aktivieren Sie die Pivot-Tabelle.
- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:
  Pivot
  Lesezeichen…
- ▶ Klicken Sie in der Liste auf den Namen des Lesezeichens.
- ▶ Klicken Sie auf Umbenennen.
- ▶ Geben Sie den neuen Namen des Lesezeichens ein.
- ▶ Klicken Sie auf OK.

Ein- und Ausblenden von Zellen

Viele Zellentypen können ausgeblendet werden:

- Dimensionsbeschriftungen,
- Kategorien, einschließlich der Zelle mit der Beschriftung und der Datenzellen in einer Zeile oder Spalte,

- Kategoriebeschriftungen (wobei die Datenzellen nicht ausgeblendet werden),
- Fußnoten, Titel und Erklärungen.

So blenden Sie Zeilen und Spalten in Tabellen aus

▶ Klicken Sie auf die Kategoriebeschriftung der auszublendenden Zeile oder Spalte, und halten Sie dabei gleichzeitig die Strg- und die Alt-Taste gedrückt.

▶ Wählen Sie die folgenden Befehle aus den Menüs für Pivot-Tabellen aus:

Ansicht
  Ausblenden

oder

▶ Klicken Sie mit der rechten Maustaste auf die Zeile oder Spalte.

▶ Wählen Sie aus dem Kontextmenü den Befehl Kategorie ausblenden aus.

So blenden Sie ausgeblendete Zeilen und Spalten in einer Tabelle wieder ein

▶ Wählen Sie eine andere Beschriftung aus derselben Dimension wie die ausgeblendete Zeile oder Spalte aus.

Wenn zum Beispiel die Kategorie *Weiblich* der Dimension "Geschlecht" ausgeblendet ist, klicken Sie auf die Kategorie *Männlich*.

▶ Wählen Sie die folgenden Befehle aus den Menüs für Pivot-Tabellen aus:

Ansicht
  Alle Kategorien in Dimensionsname einblenden

Wählen Sie zum Beispiel Alle Kategorien in Geschlecht einblenden.

oder

▶ Wählen Sie die folgenden Befehle aus den Menüs für Pivot-Tabellen aus:

Ansicht
  Alles einblenden

Dadurch werden alle ausgeblendeten Zellen in der Tabelle angezeigt. Wenn Sie aber im Dialogfeld "Tabelleneigenschaften" für diese Tabelle die Option Leere Zeilen und Spalten ausblenden ausgewählt haben, bleiben vollständig leere Zeilen oder Spalten ausgeblendet.

So können Sie die Beschriftung für die Dimension ein- bzw. ausblenden

- ▶ Aktivieren Sie die Pivot-Tabelle.
- ▶ Markieren Sie die Dimensionsbeschriftung oder eine beliebige Kategoriebeschriftung in der Dimension.
- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

  Ansicht
  Beschriftung für Dimension einblenden (oder ausblenden)

So blenden Sie Fußnoten in Tabellen ein und aus

- ▶ Wählen Sie eine Fußnote aus.
- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

  Ansicht
  Ausblenden (oder Einblenden)

So blenden Sie Erklärungen und Titel in Tabellen ein oder aus

- ▶ Wählen Sie eine Erklärung oder einen Titel aus.
- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

  Ansicht
  Ausblenden (oder Einblenden)

Bearbeiten von Ergebnissen

Das Aussehen und der Inhalt jeder Tabelle und jedes Textobjekts kann bearbeitet werden. Sie verfügen über folgende Möglichkeiten:

- Zuweisen einer Tabellenvorlage,
- Ändern der Eigenschaften der aktuellen Tabelle,
- Ändern der Eigenschaften von Zellen in der Tabelle,
- Bearbeiten von Texten,
- Hinzufügen von Fußnoten und Erklärungen zu Tabellen,
- Hinzufügen von Elementen zum Viewer,

- Kopieren und Einfügen von Ergebnissen in andere Anwendungen,
- Ändern der Breite von Datenzellen.

Ändern der Darstellung von Tabellen

Sie können die Darstellung einer Tabelle ändern, indem Sie entweder die Tabelleneigenschaften bearbeiten oder eine Tabellenvorlage zuweisen. Jede Tabellenvorlage besteht aus einer Sammlung von Tabelleneigenschaften, beispielsweise der allgemeinen Darstellung, Fußnoteneigenschaften, Zelleneigenschaften und Rahmen. Sie können eine der voreingestellten Tabellenvorlagen auswählen oder eine benutzerdefinierte Tabellenvorlage erstellen und speichern.

Tabellenvorlagen

Eine Tabellenvorlage umfaßt ein Set von Eigenschaften, welche die äußere Form einer Tabelle definieren. Sie können eine vordefinierte Tabellenvorlage auswählen oder eine eigene Tabellenvorlage erstellen.

Vor und nach dem Zuweisen von Tabellenvorlagen können Sie mithilfe der Zelleneigenschaften die Zellenformate einzelner Zellen oder Gruppen von Zellen ändern. Die bearbeiteten Zellenformate bleiben auch nach dem Zuweisen einer neuen Tabellenvorlage erhalten.

Sie können zum Beispiel einer Tabelle die Tabellenvorlage *9PUNKT* zuweisen. Wählen Sie anschließend eine Datenspalte aus, und weisen Sie dieser Spalte mit Hilfe der Registerkarte “Zellenformate” eine fette Schriftart zu. Später entscheiden Sie sich, die Tabellenvorlage *RAHMEN* auszuwählen. Die Schriftart der zuvor ausgewählten Spalte bleibt fett, während dem Rest der Tabelle die Eigenschaften für die Tabellenvorlage *RAHMEN* zugewiesen werden.

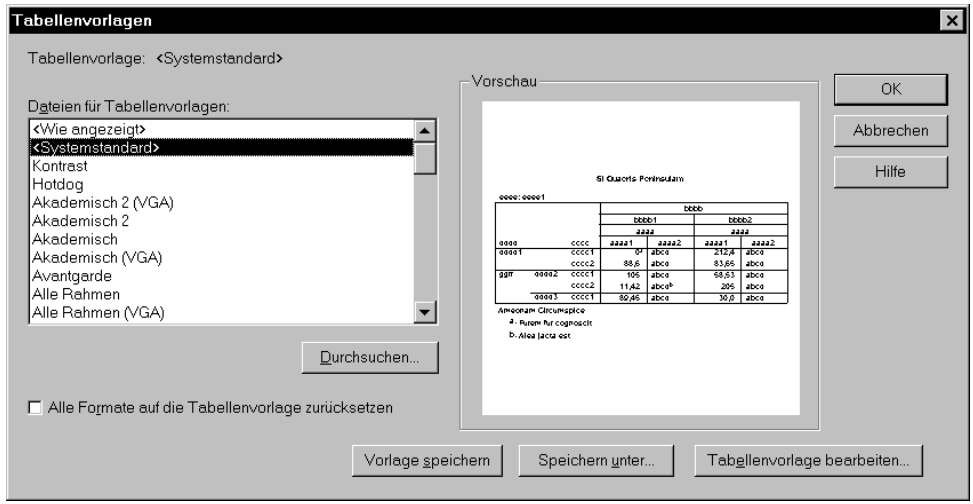
Wahlweise können Sie alle Zellen auf die Zellenformate zurücksetzen, die durch die aktuelle Tabellenvorlage definiert sind. Dadurch werden die Formate der Zellen zurückgesetzt, die zuvor bearbeitet wurden. Wenn in der Dateiliste für Tabellenvorlagen Wie angezeigt ausgewählt ist, werden alle bearbeiteten Zellen auf die aktuellen Tabelleneigenschaften zurückgesetzt.

So weisen Sie neue Tabellenvorlagen zu

- Aktivieren Sie die Pivot-Tabelle.
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Format
  Tabellenvorlagen...

  Abbildung 12-8
  Dialogfeld "Tabellenvorlagen"



- Wählen Sie eine Tabellenvorlage aus der Liste der Dateien für Tabellenvorlagen aus. Wenn Sie eine Datei in einem anderen Verzeichnis auswählen möchten, klicken Sie auf Durchsuchen.
- Klicken Sie auf OK, um der ausgewählten Pivot-Tabelle die Tabellenvorlage zuzuweisen.

So bearbeiten oder erstellen Sie Tabellenvorlagen

- Wählen Sie im Dialogfeld "Tabellenvorlagen" aus der Liste der Dateien eine Tabellenvorlage aus.
- Klicken Sie auf Tabellenvorlage bearbeiten.

- Passen Sie die Tabelleneigenschaften an die gewünschten Attribute an, und klicken Sie anschließend auf OK.
- Klicken Sie auf Vorlage speichern, um die bearbeitete Tabellenvorlage zu speichern, oder klicken Sie auf Speichern unter, um sie als neue Tabellenvorlage zu speichern.

  Änderungen in einer Tabellenvorlage wirken sich nur auf die ausgewählte Pivot-Tabelle aus. Bearbeitete Tabellenvorlagen werden anderen Tabellen, die diese Tabellenvorlage verwenden, nicht automatisch zugewiesen. Hierfür müßten Sie die betreffenden Tabellen auswählen und die Tabellenvorlage erneut zuweisen.

Tabelleneigenschaften

Im Dialogfeld "Tabelleneigenschaften" können Sie die allgemeinen Eigenschaften einer Tabelle festlegen, Zelleneigenschaften für verschiedene Teile einer Tabelle bestimmen und diese Eigenschaften als Tabellenvorlage speichern. In den Registerkarten dieses Dialogfelds können Sie folgendes tun:

- Festlegen allgemeiner Eigenschaften, beispielsweise das Ausblenden leerer Zeilen und Spalten und das Anpassen der Druckeigenschaften,
- Festlegen des Formats und der Position von Fußnotenzeichen,
- Festlegen spezieller Formate für Zellen im Datenbereich, für Zeilen- und Spaltenbeschriftungen und für andere Bereiche der Tabelle,
- Festlegen der Breite und Farbe für die Rahmenlinien der einzelnen Tabellenbereiche.

So ändern Sie die Eigenschaften von Pivot-Tabellen

- Doppelklicken Sie auf eine beliebige Stelle in der Pivot-Tabelle, um diese zu aktivieren.
- Wählen Sie die folgenden Befehle aus den Menüs für Pivot-Tabellen aus:

  Format
  Tabelleneigenschaften…
- Wählen Sie eine Registerkarte aus (Allgemein, Fußnoten, Zellenformate, Rahmen oder Drucken).
- Wählen Sie die gewünschten Optionen aus.

▶ Klicken Sie auf OK oder Zuweisen.

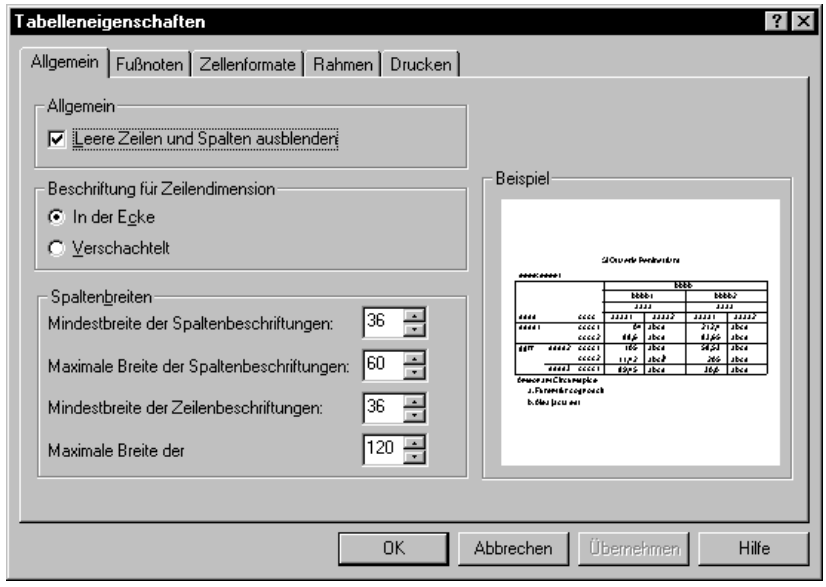
Die neuen Eigenschaften werden der ausgewählten Pivot-Tabelle zugewiesen. Wenn Sie die neuen Tabelleneigenschaften nicht nur der ausgewählten Tabelle, sondern einer Tabellenvorlage hinzufügen möchten, bearbeiten Sie die Tabellenvorlage. Wählen Sie dazu im Menü “Format” den Befehl “Tabellenvorlage”.

Tabelleneigenschaften: Allgemein

Einige Eigenschaften betreffen die ganze Tabelle. Sie verfügen über folgende Möglichkeiten:

- Ein- oder Ausblenden von leeren Zeilen und Spalten. (Bei einer leeren Zeile oder Spalte steht in keiner der Datenzellen ein Wert.)
- Festlegen der Platzierung der Zeilenbeschriftungen, die sich in der oberen linken Ecke befinden oder verschachtelt sein können.
- Festlegen der maximalen und minimalen Spaltenbreite (angegeben in Punkt).

Abbildung 12-9
Dialogfeld “Tabelleneigenschaften”, Registerkarte “Allgemein”



So ändern Sie allgemeine Tabelleneigenschaften

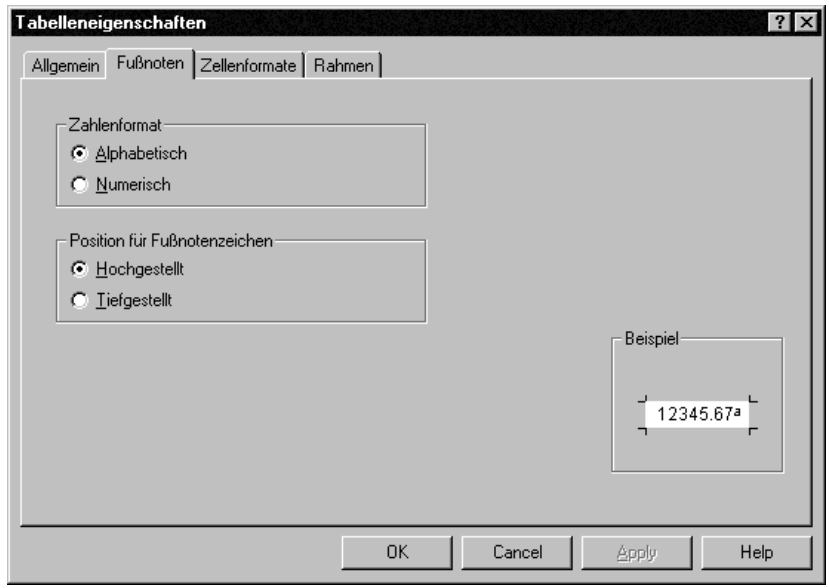
- ▶ Klicken Sie auf die Registerkarte Allgemein.
- ▶ Wählen Sie die gewünschten Optionen aus.
- ▶ Klicken Sie auf OK oder Zuweisen.

Tabelleneigenschaften: Fußnoten

Zu den Eigenschaften von Fußnotenzeichen gehören zum Beispiel Format und Position in bezug auf den Text.

- Als Format für Fußnotenzeichen können Sie entweder Ziffern (1, 2, 3, ...) oder Buchstaben (a, b, c, ...) festlegen.
- Die Fußnotenzeichen können hochgestellt oder tiefgestellt eingefügt werden.

Abbildung 12-10
Dialogfeld "Tabelleneigenschaften", Registerkarte "Fußnoten"



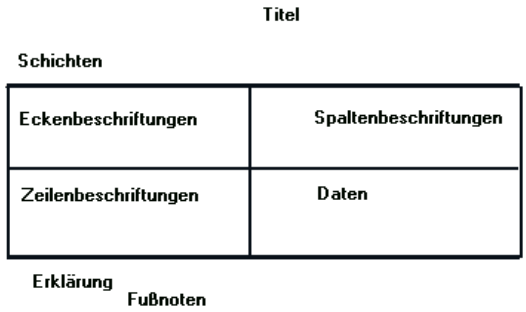
So ändern Sie die Eigenschaften von Fußnotenzeichen

- Klicken Sie auf die Registerkarte Fußnoten.
- Wählen Sie ein Zahlenformat für die Fußnoten aus.
- Wählen Sie eine Position für Fußnotenzeichen aus.
- Klicken Sie auf OK oder Zuweisen.

Tabelleneigenschaften: Zellenformate

In bezug auf die Formatierung ist eine Tabelle in verschiedene Bereiche aufgeteilt: Titel, Schichten, Eckenbeschriftungen, Zeilenbeschriftungen, Spaltenbeschriftungen, Daten, Erklärungen und Fußnoten. Für jeden Bereich der Tabelle können Sie die entsprechenden Zellenformate ändern. Folgende Zellenformate können bearbeitet werden: Texteigenschaften (z. B.Schriftart, Größe, Farbe und Schnitt), horizontale und vertikale Ausrichtung, Zellenschattierung, Vordergrund und Hintergrundfarben sowie die inneren Zellenränder.

Abbildung 12-11
Bereiche einer Tabelle



Zellenformate werden immer ganzen Bereichen (Informationskategorien) zugewiesen. Sie sind nicht Eigenschaften einzelner Zellen. Diese Unterscheidung ist besonders beim Pivotieren von Tabellen wichtig.

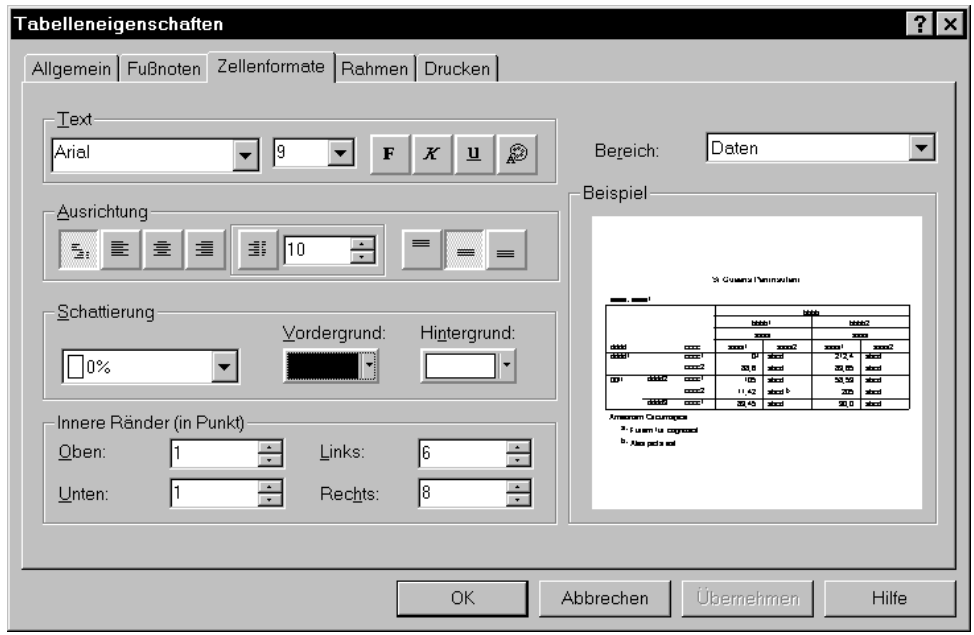
Ein Beispiel:

- Wenn Sie als Zellenformat für Spaltenbeschriftungen eine fette Schriftart festlegen, werden die Spaltenbeschriftungen fett angezeigt, und zwar unabhhängig davon, was gerade in der Spaltendimension angezeigt wird. Wenn

Sie ein Element aus der Spaltendimension in eine andere Dimension verschieben, bleibt die fette Formatierung für dieses Element nicht erhalten.

- Wenn Sie hingegen die Spaltenbeschriftungen fett formatieren, indem Sie die Zellen in einer aktivierten Pivot-Tabelle markieren und auf der Symbolleiste auf die Schaltfläche "Fett" klicken, bleibt der Inhalt dieser Zellen auch beim Verschieben in andere Dimensionen immer fett formatiert. Die Spaltenbeschriftungen behalten diese Formatierung dann nicht für andere Elemente, die in die Spaltendimension verschoben werden.

Abbildung 12-12
Dialogfeld "Tabelleneigenschaften", Registerkarte "Zellenformate"



So ändern Sie Zellenformate

- ▶ Klicken Sie auf die Registerkarte Zellenformate.

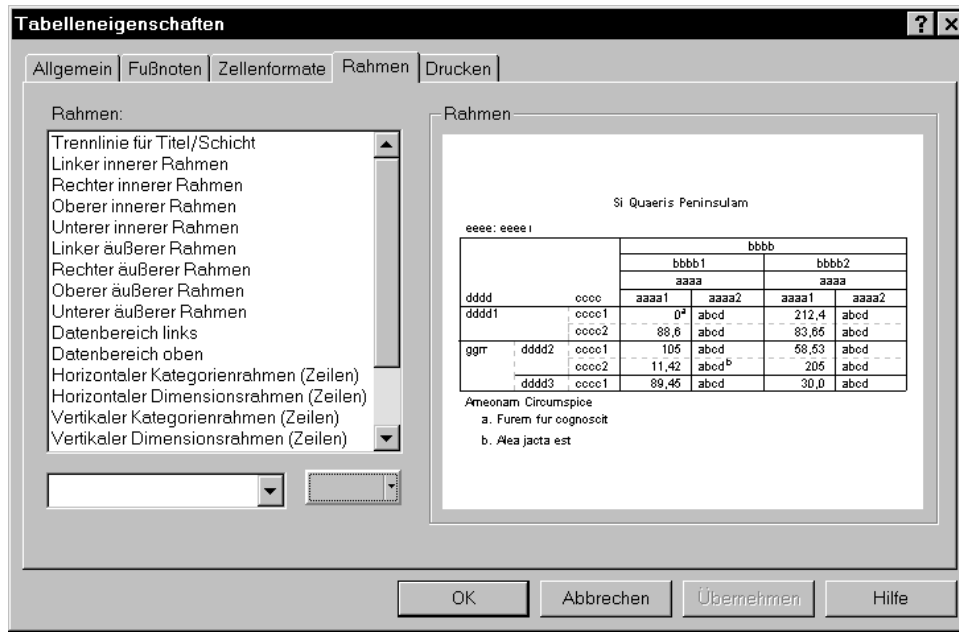
- ▶ Wählen Sie einen Bereich aus der Dropdown-Liste aus, oder klicken Sie auf einen Bereich in der Vorschau.

- Wählen Sie die Eigenschaften für diesen Bereich aus. Ihre Auswahl wird in der Vorschau angezeigt.
- Klicken Sie auf OK oder Zuweisen.

Tabelleneigenschaften: Rahmen

Sie können für jeden Rahmen in einer Tabelle einen Linienstil und eine Farbe auswählen. Wenn Sie Kein in der Liste für die Linienstile auswählen, wird an der entsprechenden Position keine Linie gezeichnet.

Abbildung 12-13
Dialogfeld "Tabelleneigenschaften", Registerkarte "Rahmen"



So ändern Sie Rahmen in Tabellen

- Klicken Sie auf die Registerkarte Rahmen.
- Wählen Sie eine Rahmenposition aus. Klicken Sie hierfür auf den entsprechenden Namen in der Liste, oder klicken Sie auf eine Linie in der Vorschau. (Halten Sie

für die Auswahl mehrerer Namen beim Klicken mit der Maus gleichzeitig die Umschalt-Taste gedrückt. Bei nicht aufeinanderfolgenden Namen halten Sie statt dessen die Strg-Taste gedrückt, während Sie auf die einzelnen Namen klicken.)

- Wählen Sie einen Linienstil, oder wählen Sie die Option Kein aus.
- Wählen Sie eine Farbe aus.
- Klicken Sie auf OK oder Zuweisen.

So lassen Sie die ausgeblendeten Rahmen in einer Pivot-Tabelle anzeigen

Bei Tabellen, bei denen nur wenige Rahmen sichtbar sind, können Sie die ausgeblendeten Rahmen anzeigen lassen. Dadurch werden Vorgänge wie das Ändern von Spaltenbreiten vereinfacht. Die ausgeblendeten Rahmen (Gitterlinien) werden im Viewer angezeigt, jedoch nicht gedruckt.

- Doppelklicken Sie auf eine beliebige Stelle in der Pivot-Tabelle, um diese zu aktivieren.
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Ansicht
  Gitterlinien

Tabelleneigenschaften: Drucken

Sie können folgende Eigenschaften für den Ausdruck von Pivot-Tabellen beeinflussen:

- Drucken aller Schichten oder nur der obersten Schicht der Tabelle und Drucken jeder Schicht auf einer getrennten Seite. (Diese Einstellung beeinflußt nur den Ausdruck, nicht jedoch die Anzeige von Schichten im Viewer.)
- Horizontales oder vertikales Verkleinern einer Tabelle zur Anpassung an die Seitengröße beim Drucken.
- Legen Sie die Einstellungen für Witwen/Waisen-Zeilen fest, in dem Sie die Mindestanzahl von Zeilen und Spalten festlegen, die in einem beliebigen gedruckten Abschnitt einer Tabelle enthalten sind, wenn die Tabelle zu breit und/oder zu lang für die definierte Seitengröße ist. (*Anmerkung*: Wenn eine Tabelle nicht mehr auf die aktuelle Seite passt, weil sich bereits andere Ausgaben darüber befinden, die Tabelle im Prinzip aber auf eine Seite passen würde, wird

sie automatisch auf einer neuen Seite gedruckt. Dies geschieht unabhängig von den Einstellungen für Witwen/Waisen-Zeilen.)

- Berücksichtigen von Fortsetzungstexten für Tabellen, die nicht auf eine einzige Seite passen. Fortsetzungstexte können am unteren oder oberen Rand jeder Seite angezeigt werden. Wenn Sie keine dieser Optionen auswählen, werden Fortsetzungstexte nicht angezeigt.

So wählen Sie Optionen für das Drucken von Pivot-Tabellen aus

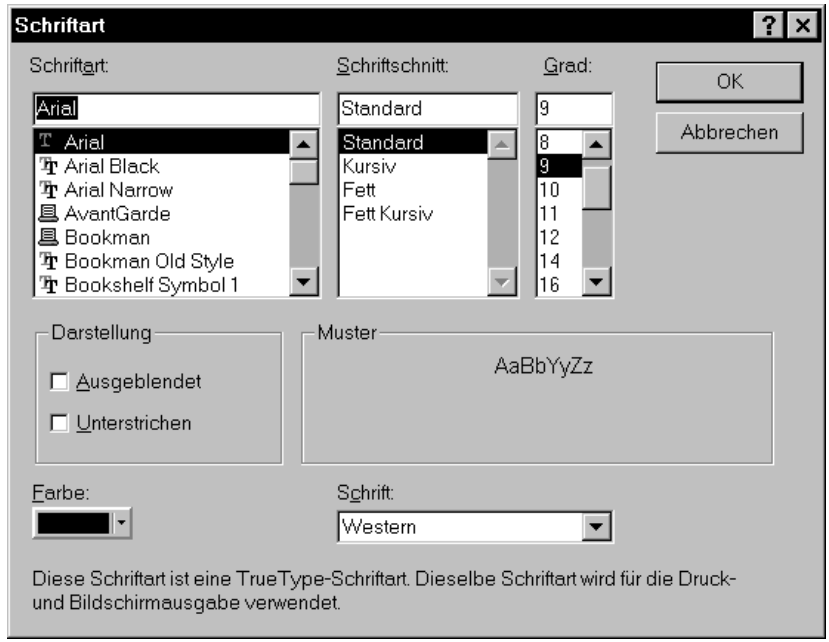
- Klicken Sie auf die Registerkarte Drucken.
- Wählen Sie die gewünschten Druckoptionen aus.
- Klicken Sie auf OK oder Zuweisen.

Schriftart

Mit Tabellenvorlagen können Sie Merkmale von Schriftarten für verschiedene Bereiche einer Tabelle festlegen. Außerdem können Sie die Schriftart für jede einzelne Zelle getrennt festlegen. Sie können beispielsweise Schriftart, Schriftschnitt, Schriftgröße, Schriftfarbe für den Text in einer Zelle festlegen. Sie können den Text auch ausblenden oder unterstreichen.

Eigenschaften von Schriftarten, die Sie für eine Zelle festlegen, wirken sich auf alle Tabellenschichten aus, in denen dieselbe Zelle vorkommt.

Abbildung 12-14
Dialogfeld "Schriftart"



So ändern Sie die Schriftart in einer Zelle

- ▶ Aktivieren Sie die Pivot-Tabelle, und markieren Sie den gewünschten Text.
- ▶ Wählen Sie die folgenden Befehle aus den Menüs für Pivot-Tabellen aus:

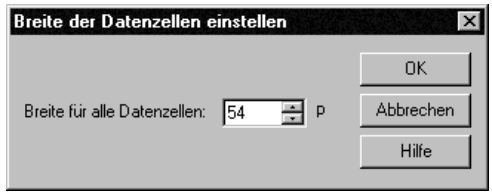
  Format
  Schriftart...

  Sie können optional die Schriftart, den Schriftschnitt, den Schriftgrad in Punkt und die Farbe des Textes festlegen sowie ein Sprachskript auswählen. Außerdem können Sie angeben, ob der Text ausgeblendet oder unterstrichen werden soll.

Breite der Datenzellen

Mit "Breite der Datenzellen" können Sie für alle Datenzellen dieselbe Breite festlegen.

Abbildung 12-15
Dialogfeld "Breite der Datenzellen einstellen"



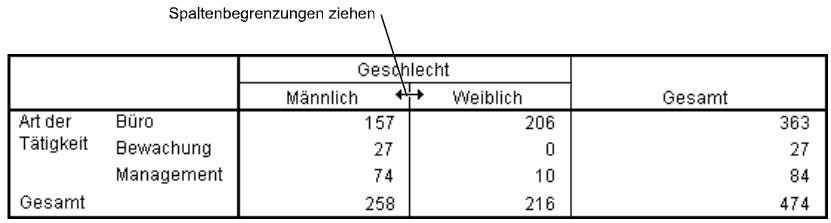
So ändern Sie die Breite der Datenzellen

- Aktivieren Sie die Pivot-Tabelle.
- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Format
  Breite der Datenzellen...
- Geben Sie einen Wert für die Zellenbreite an.

So ändern Sie die Breite von Spalten in Pivot-Tabellen

- Doppelklicken Sie auf eine beliebige Stelle in der Tabelle, um diese zu aktivieren.
- Zeigen Sie mit dem Mauszeiger auf den rechten Rand der zu ändernden Spalte mit Kategoriebeschriftungen. (Der Mauszeiger nimmt die Form eines Doppelpfeils an.)
- Ziehen Sie den Rahmenrand mit gedrückter Maustaste auf die neue Position.

Abbildung 12-16
Ändern der Breite einer Spalte



Sie können die vertikalen Rahmen für Kategorien und Dimensionen im Bereich der Zeilenbeschriftungen ändern, unabhängig davon, ob sie gerade angezeigt werden.

- Ziehen Sie den Mauszeiger über die Zeilenbeschriftungen, bis er als Doppelpfeil dargestellt wird.
- Ziehen Sie den Rahmenrand auf die neue Breite.

Zelleneigenschaften

Zelleneigenschaften werden ausgewählten Zellen zugewiesen. Sie können das Werteformat, die Ausrichtung, die Ränder und die Schattierung ändern. Zelleneigenschaften setzen Tabelleneigenschaften außer Kraft. Wenn Sie also die Tabelleneigenschaften ändern, bleiben vorher zugewiesene Zelleneigenschaften erhalten.

So ändern Sie Zelleneigenschaften

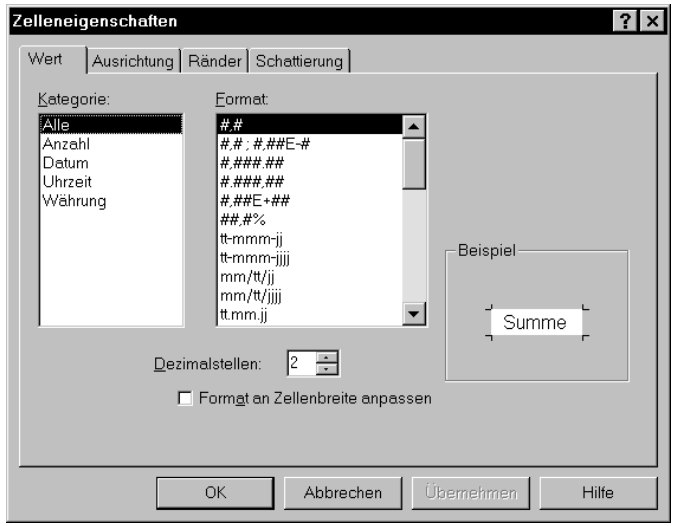
- Aktivieren Sie eine Tabelle, und wählen Sie eine Zelle in der Tabelle aus.
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Format
  Zelleneigenschaften…

Zelleneigenschaften: Wert

Auf dieser Registerkarte können Sie das Werteformat für eine Zelle festlegen. Sie können die Formate für Zahlen, Datum, Zeit und Währung bestimmen und die Anzahl der angezeigten Dezimalstellen festlegen.

Abbildung 12-17
Dialogfeld "Zelleneigenschaften", Registerkarte "Wert"



So ändern Sie Werteformate in Zellen

- Klicken Sie auf die Registerkarte Wert.
- Wählen Sie eine Kategorie und ein Format aus.
- Wählen Sie die Anzahl von Dezimalstellen aus.

So ändern Sie Werteformate für Spalten in Pivot-Tabellen

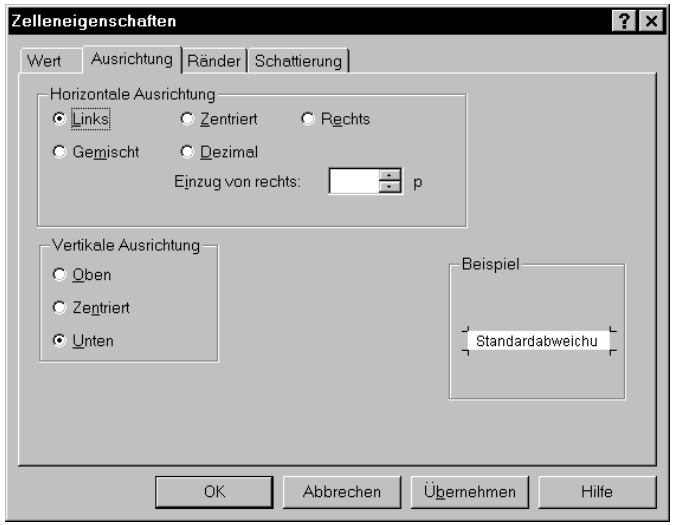
- Klicken Sie auf die Spaltenbeschriftung, und halten Sie dabei gleichzeitig die Strg- und die Alt-Taste gedrückt.
- Klicken Sie mit der rechten Maustaste auf die markierte Spalte.
- Wählen Sie aus dem angezeigten Kontextmenü den Befehl Zelleneigenschaften.
- Klicken Sie auf die Registerkarte Wert.
- Wählen Sie das Format, das Sie der Spalte zuweisen möchten.

Sie können diese Methode verwenden, um Prozent- und Dollarzeichen hinzuzufügen oder zu unterdrücken, die Anzahl der angezeigten Dezimalstellen zu ändern sowie zwischen der wissenschaftlichen Notation und der regulären numerischen Anzeige umzuschalten.

Zelleneigenschaften: Ausrichtung

Auf dieser Registerkarte können Sie die horizontale und vertikale Ausrichtung des Textes für eine Zelle festlegen. Wenn Sie die Option "Gemischt" auswählen, wird der Inhalt der Zelle entsprechend dem Typ des Inhalts ausgerichtet (Zahl, Datum oder Text).

Abbildung 12-18
Dialogfeld "Zelleneigenschaften", Registerkarte "Ausrichtung"



So ändern Sie die Ausrichtung in Zellen

- Wählen Sie in der Tabelle eine Zelle aus.
- Wählen Sie die folgenden Befehle aus den Menüs für Pivot-Tabellen aus:

  Format
  Zelleneigenschaften…

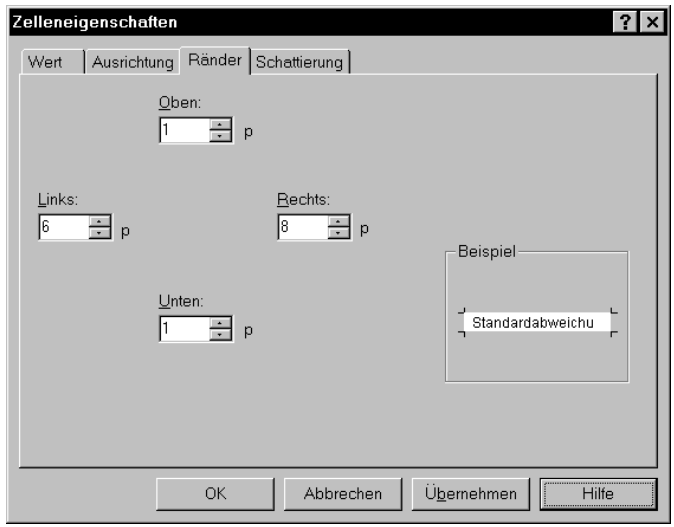
- Klicken Sie auf die Registerkarte Ausrichtung.

  Sie können die ausgewählten Eigenschaften für die Ausrichtung der Zelle direkt in der Vorschau betrachten.

Zelleneigenschaften: Ränder

Auf dieser Registerkarte können Sie die Einstellungen für die Ränder einer Zelle festlegen.

Abbildung 12-19
Dialogfeld "Zelleneigenschaften", Registerkarte "Ränder"



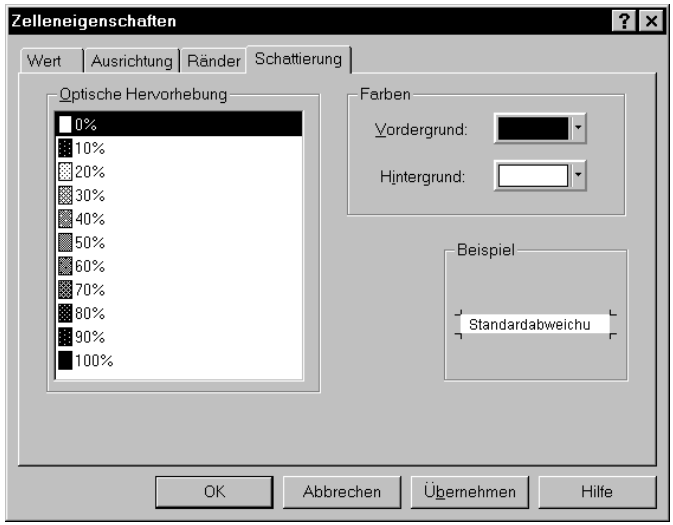
So ändern Sie Ränder von Zellen

- Klicken Sie auf die Registerkarte Ränder.
- Wählen Sie die Einstellungen für die vier Ränder.

Zelleneigenschaften: Schattierung

Auf dieser Registerkarte können Sie die Schattierung sowie Vordergrund- und Hintergrundfarben für einen ausgewählten Zellenbereich festlegen. Die Farbe des Textes wird durch diese Einstellung nicht geändert.

Abbildung 12-20
Dialogfeld "Zelleneigenschaften", Registerkarte "Schattierung"



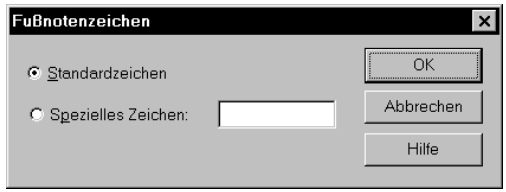
So ändern Sie die Schattierung in Zellen

- Klicken Sie auf die Registerkarte Schattierung.
- Wählen Sie die Hervorhebung und Farben für die Zelle aus.

Fußnotenzeichen

Mit dem Befehl "Fußnotenzeichen" ändern Sie die Zeichen, die zum Kennzeichnen von Fußnoten verwendet werden.

Abbildung 12-21
Dialogfeld "Fußnotenzeichen"



So ändern Sie die Zeichen für Fußnotenzeichen

- Wählen Sie eine Fußnote aus.
- Wählen Sie die folgenden Befehle aus den Menüs für Pivot-Tabellen aus:

  Format
  Fußnotenzeichen…
- Geben Sie ein oder zwei Zeichen ein.

So numerieren Sie Fußnoten neu

Nach dem Pivotieren einer Tabelle durch Verschieben von Zeilen, Spalten und Schichten ist die Reihenfolge der Fußnoten unter Umständen durcheinander. So numerieren Sie die Fußnoten neu:

- Aktivieren Sie die Pivot-Tabelle.
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Format
  Fußnoten neu numerieren…

Auswählen von Zeilen und Spalten in Pivot-Tabellen

In Pivot-Tabellen gibt es einige Beschränkungen beim Auswählen ganzer Zeilen oder Spalten. Die optische Markierung, welche die markierte Zeile oder Spalte kennzeichnet, kann gegebenenfalls nicht zusammenhängende Bereiche der Tabelle umfassen.

So wählen Sie eine Zeile oder Spalte in einer Pivot-Tabelle aus

- Doppelklicken Sie auf eine beliebige Stelle in der Pivot-Tabelle, um diese zu aktivieren.
- Klicken Sie auf eine Zeilen- oder Spaltenbeschriftung.
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Auswählen
  Datenzellen und Beschriftung

  oder

- Klicken Sie auf die Zeilen- oder Spaltenbeschriftung und halten Sie gleichzeitig die Strg- und die Alt-Taste gedrückt.

  Wenn die Tabelle mehr als eine Dimension im Zeilen- oder Spaltenbereich enthält, kann es vorkommen, daß die hervorgehobene Auswahl nicht zusammenhängende Zellen umfaßt.

Anpassen von Ergebnissen in Pivot-Tabellen

Viele Elemente im Viewer enthalten Texte. Sie können diese Texte bearbeiten und neue Texte hinzufügen. So können Sie Pivot-Tabellen anpassen:

- Bearbeiten von Texten in Pivot-Tabellen-Zellen,
- Hinzufügen von Erklärungen und Fußnoten.

So passen Sie Texte in Zellen an

- Aktivieren Sie die Pivot-Tabelle.
- Doppelklicken Sie auf die Zelle.
- Bearbeiten Sie den Text.
- Drücken Sie die Eingabetaste zum Aufzeichnen der Änderungen, oder drücken Sie die Esc-Taste, um den vorherigen Inhalt der Zelle wiederherzustellen.

So fügen Sie Tabellen Erklärungen hinzu

▶ Wählen Sie die folgenden Befehle aus den Menüs für Pivot-Tabellen aus:

Einfügen
Erklärung

Daraufhin wird das Wort Erklärung am unteren Rand der Tabelle angezeigt.

▶ Doppelklicken Sie auf den Text Erklärung, und ersetzen Sie diesen durch Ihre eigene Erkärung.

So fügen Sie Tabellen Fußnoten hinzu

Sie können jedem Element einer Tabelle eine Fußnote hinzufügen.

▶ Klicken Sie auf einen Titel, eine Zelle oder eine Erklärung in einer aktivierten Pivot-Tabelle.

▶ Wählen Sie die folgenden Befehle aus den Menüs für Pivot-Tabellen aus:

Einfügen
Fußnote...

▶ Doppelklicken Sie auf das Wort Fußnote, und ersetzen Sie dieses durch den Fußnotentext.

Drucken von Pivot-Tabellen

Das Aussehen gedruckter Pivot-Tabellen wird von mehreren Faktoren bestimmt. Diese Faktoren können Sie durch Ändern der Attribute für Pivot-Tabellen beeinflussen.

- Bei multidimensionalen Pivot-Tabellen (Tabellen mit Schichten) können Sie entweder alle Schichten oder nur die oberste (sichtbare) Schicht ausdrucken.
- Sie können zu lange oder zu breite Pivot-Tabellen automatisch an die Seitengröße anpassen. Sie können auch die Positionen von Tabellenumbrüchen und Seitenumbrüchen festlegen.

Verwenden Sie den Befehl “Seitenansicht” im Menü “Datei”, um sich anzeigen zu lassen, wie die Pivot-Tabellen auf der gedruckten Seite aussehen würden.

So drucken Sie die verborgenen Schichten von Pivot-Tabellen

- ▶ Doppelklicken Sie auf eine beliebige Stelle in der Tabelle, um diese zu aktivieren.
- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:
  Format
  Tabelleneigenschaften…
- ▶ Wählen Sie auf der Registerkarte "Drucken" die Option Alle Schichten drucken.

  Sie können auch jede Schicht einer Pivot-Tabelle auf einer separaten Seite ausdrucken.

Festlegen von Tabellenumbrüchen für breite und lange Tabellen

Pivot-Tabellen, die zu breit oder zu lang sind, um innerhalb der definierten Seitengröße gedruckt zu werden, werden automatisch aufgeteilt und in mehreren Abschnitten gedruckt. Bei breiten Tabellen werden mehrere Abschnitte auf derselben Seite gedruckt, wenn genügend Platz zur Verfügung steht. Sie verfügen über folgende Möglichkeiten:

- Festlegen der Zeilen und Spalten, an denen große Tabellen geteilt werden,
- Angeben der Zeilen und Spalten, bei denen Tabellen nicht getrennt werden sollen,
- Anpassen von umfangreichen Tabellen an die definierte Seitengröße.

So legen Sie Zeilen- und Spaltenumbrüche für gedruckte Pivot-Tabellen fest

- ▶ Aktivieren Sie die Pivot-Tabelle.
- ▶ Klicken Sie auf die Spaltenbeschriftung links neben der Stelle, oder klicken Sie auf die Zeilenbeschriftung über der Stelle, an der Sie den Umbruch einfügen möchten.
- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:
  Format
  Umbruch hier

So legen Sie Zeilen oder Spalten fest, die nicht getrennt werden sollen

- ▶ Aktivieren Sie die Pivot-Tabelle.

- Wählen Sie die Beschriftungen der Zeilen oder Spalten aus, die nicht getrennt werden sollen. Sie können mehrere Zeilen- oder Spaltenbeschriftungen auswählen, indem Sie mit gedrückter Maustaste ziehen oder die Umschalttaste gedrückt halten und auf die erste und die letzte auszuwählende Beschriftung klicken.

- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Format
  Zusammenhalten

So passen Sie Pivot-Tabellen an die Seitengröße an

- Aktivieren Sie die Pivot-Tabelle.

- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Format
  Tabelleneigenschaften…

- Klicken Sie auf die Registerkarte Drucken.

- Klicken Sie auf Breite Tabelle auf Seitengröße verkleinern.

  und/oder

- Klicken Sie auf Lange Tabelle auf Seitengröße verkleinern.

Kapitel
13

Arbeiten mit der Befehlssyntax

SPSS bietet eine leistungsfähige Befehlssprache, mit deren Hilfe Sie viele häufig durchzuführende Aufgaben speichern und automatisieren können. Sie bietet außerdem einige Funktionen, die nicht über die Menüs und Dialogfelder zur Verfügung stehen.

Auf die meisten SPSS-Befehle können Sie über die Menüs und Dialogfelder zugreifen. Einige Befehle und Optionen sind aber nur in der SPSS-Befehlssprache verfügbar. Mit der Befehlssprache verfügen Sie außerdem über die Möglichkeit, Jobs in einer Syntaxdatei zu speichern. Sie können eine Analyse dann zu einem späteren Zeitpunkt wiederholen oder diese automatisch mit dem Produktionsmodus ausführen lassen.

Eine Syntaxdatei ist eine einfache Textdatei, die SPSS-Befehle enthält. Es ist zwar möglich, ein Syntax-Fenster zu öffnen und Befehle einzugeben. Mit den folgenden Funktionen kann SPSS Sie jedoch beim Erstellen einer Syntaxdatei unterstützen:

- Einfügen von Syntaxbefehlen aus Dialogfeldern
- Kopieren der Syntax aus dem Ausgabe-Log
- Kopieren der Syntax aus der Journaldatei

Detaillierte Informationen zur Befehlssyntax sind auf zwei Arten verfügbar: als Bestandteil des umfassenden Hilfesystems und als separates Dokument im PDF-Format in der *SPSS Command Syntax Reference*, die auch über das Menü "Hilfe" verfügbar ist.

Sie können auf kontextsensitive Hilfe für den aktuellen Befehl in einem Syntax-Fenster zugreifen, indem Sie die F1-Taste drücken.

Regeln für die Syntax

Wenn Sie während einer SPSS-Sitzung Befehle in einem Befehlssyntax-Fenster ausführen, werden diese Befehle im **interaktiven** Modus ausgeführt.

Die folgenden Regeln gelten für die Angaben von Befehlen im interaktiven Modus:

- Jeder Befehl muss mit einem Punkt abgeschlossen werden. Es empfiehlt sich jedoch, den Befehlsabschluss bei `BEGIN DATA` wegzulassen, sodass in der Zeile enthaltene Daten als eine fortlaufende Angabe behandelt werden.
- Der Befehlsabschluss muss das letzte nichtleere Zeichen in einem Befehl sein.
- Falls kein Punkt als Befehlsabschluss vorhanden ist, wird eine leere Zeile als Befehlsabschluss interpretiert.
- Befehle können in jeder Spalte einer Zeile beginnen und für beliebig viele Zeilen fortgesetzt werden. Die einzige Ausnahme ist der Befehl `END DATA`, der in der ersten Spalte der ersten Zeile nach dem Ende der Daten beginnen muss.

Anmerkung: Um die Kompatibilität mit anderen Modi für die Ausführung von Befehlen (einschließlich Befehlsdateien, die mit den Befehlen `INSERT` oder `INCLUDE` in einer aktiven Sitzung ausgeführt werden) zu wahren, darf keine Zeile der Befehlssyntax 256 Byte überschreiten.

- Die meisten Unterbefehle werden durch Schrägstriche (/) voneinander getrennt. Der Schrägstrich vor dem ersten Unterbefehl ist in der Regel optional.
- Variablennamen müssen vollständig ausgeschrieben werden.
- Text in Apostrophen oder Anführungszeichen muß sich auf einer Zeile befinden.
- Zum Kennzeichnen der Dezimalstellen muss unabhängig von den regionalen oder Gebietsschemaeinstellungen der Punkt (.) verwendet werden.
- Variablenamen, die mit einem Punkt enden, können bei Befehlen, die aus einem Dialogfeld übernommen wurden, Fehler hervorrufen. Sie dürfen solche Variablennamen nicht in Dialogfeldern verwenden und sollten sie auch generell vermeiden.

Bei der SPSS-Befehlssyntax wird nicht zwischen Groß- und Kleinschreibung unterschieden. Für viele Befehle können Abkürzungen aus drei oder vier Zeichen verwendet werden. Sie können beliebig viele Zeilen zur Angabe eines einzelnen Befehls verwenden. An fast jedem Punkt, an dem ein Leerzeichen zulässig ist, können Sie beliebig viele Leerzeilen oder Zeilenumbrüche einfügen, beispielsweise bei Schrägstrichen, runden Klammern, arithmetischen Operatoren oder zwischen Variablennamen. Beispiel:

```
FREQUENCIES
   VARIABLES=TÄTIG GESCHL
```

```
/PERCENTILES=25 50 75
/BARCHART.
```

und

```
freq var=tätig geschl /percent=25 50 75 /bar.
```

Mit beiden Formen wird das gleiche Ergebnis erzielt.

INCLUDE-Dateien

Befehlsdateien, die mit dem Befehl `INCLUDE` ausgeführt werden, unterliegen den Syntaxregeln für den Stapelverarbeitungsmodus.

Die folgenden Regeln gelten für die Angabe von Befehlen im Stapelverarbeitungs- und Produktionsmodus:

- Alle Befehle in der Befehlsdatei müssen in Spalte 1 beginnen. Sie können in der ersten Spalte Plus- (+) oder Minuszeichen (-) verwenden, wenn Sie Befehle einrücken möchten, um die Befehlsdatei lesbarer zu machen.
- Wenn für einen Befehl mehrere Zeilen verwendet werden, muss Spalte 1 jeder folgenden Zeile leer sein.
- Befehlsabschlüsse sind optional.
- Eine Zeile darf 256 Byte nicht überschreiten. Jedes zusätzliches Zeichen wird abgeschnitten.

Falls noch keine Befehlsdateien vorliegen, in denen der Befehl `INCLUDE` bereits verwendet wird, sollten Sie statt dessen eher den Befehl `INSERT` verwenden, weil hiermit Befehlsdateien berücksichtigt werden, die beiden Regelsätzen entsprechen. Wenn Sie die Befehlssyntax mit der Schaltfläche “Einfügen” aus einem Dialogfeld in das Syntax-Fenster übernehmen, ist das Format für alle Betriebsmodi geeignet. Weitere Informationen finden Sie in der *SPSS Command Syntax Reference* (im PDF-Format über das Menü “Hilfe” verfügbar).

Übernehmen der Befehlssyntax aus Dialogfeldern

Am einfachsten erstellen Sie eine Befehlssyntax-Datei, indem Sie die entsprechenden Optionen in einem SPSS-Dialogfeld auswählen und die Syntax für diese Auswahl in ein Syntax-Fenster übernehmen. Sie können eine Job-Datei erstellen, indem Sie die Syntax einer längeren Analyse Schritt für Schritt übernehmen. Mit dieser Datei können Sie die Analyse zu einem späteren Zeitpunkt wiederholen oder einen Job mit dem SPSS-Produktionsmodus ausführen lassen.

In einem Syntax-Fenster können Sie die übernommene Syntax ausführen, bearbeiten und in einer Syntaxdatei speichern.

So übernehmen Sie die Befehlssyntax aus Dialogfeldern:

- ▶ Öffnen Sie das Dialogfeld, und treffen Sie die gewünschte Auswahl.
- ▶ Klicken Sie auf Einfügen.

Die Befehlssyntax wird in das Haupt-Syntax-Fenster eingefügt. Falls kein Syntax-Fenster geöffnet ist, öffnet SPSS eines neues Syntax-Fenster und fügt die Syntax dort ein.

Abbildung 13-1
Aus einem Dialogfeld eingefügte Befehlssyntax

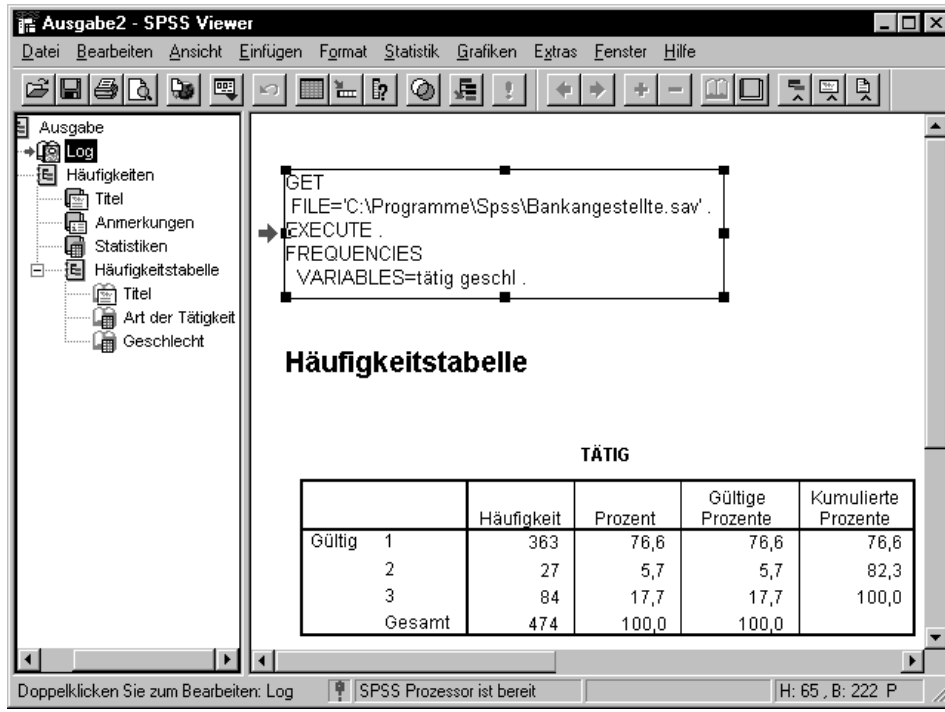
Anmerkung: Wenn Sie ein Dialogfeld über die Menüs in einem Skriptfenster aufrufen, wird der Code für das Ausführen einer Syntax über ein Skript in das Skriptfenster eingefügt.

Kopieren von Syntax aus dem Ausgabe-Log

Sie können eine Syntaxdatei erstellen, indem Sie die Befehlssyntax aus dem SPSS-Log kopieren, das im Viewer angezeigt wird. Dazu müssen Sie in den Viewer-Einstellungen (Menü “Bearbeiten”, “Optionen”, Registerkarte “Viewer”) die Option Befehle im Log anzeigen auswählen, bevor Sie die Analyse ausführen. Jeder Befehl wird dann zusammen mit der Ausgabe der Analyse im Viewer angezeigt.

In einem Syntax-Fenster können Sie die übernommene Syntax ausführen, bearbeiten und in einer Syntaxdatei speichern.

Abbildung 13-2
Befehlssyntax im Log



So kopieren Sie die Syntax aus dem Ausgabe-Log

- Bevor Sie die Analyse ausführen, wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Optionen...

- Wählen Sie auf der Registerkarte "Viewer" die Option Befehle im Log anzeigen.

  Wenn Sie die Analyse ausführen, werden die SPSS-Befehle für Ihre Auswahl im Dialogfeld im Log aufgezeichnet.

- Öffnen Sie eine vorher gespeicherte Syntaxdatei, oder erstellen Sie eine neue. Wählen Sie zum Erstellen einer neuen Syntaxdatei die folgenden Befehle aus den Menüs aus:

  Datei
  Neu
  Syntax

- Doppelklicken Sie im Viewer auf einen Log-Eintrag, um diesen zu aktivieren.

- Klicken Sie, und ziehen Sie mit der Maus, um die Syntax zu markieren, die Sie kopieren möchten.

- Wählen Sie die folgenden Befehle aus den Menüs des Viewers aus:

  Bearbeiten
  Kopieren

- Wählen Sie die folgenden Befehle aus den Menüs in einem Syntax-Fenster aus:

  Bearbeiten
  Einfügen

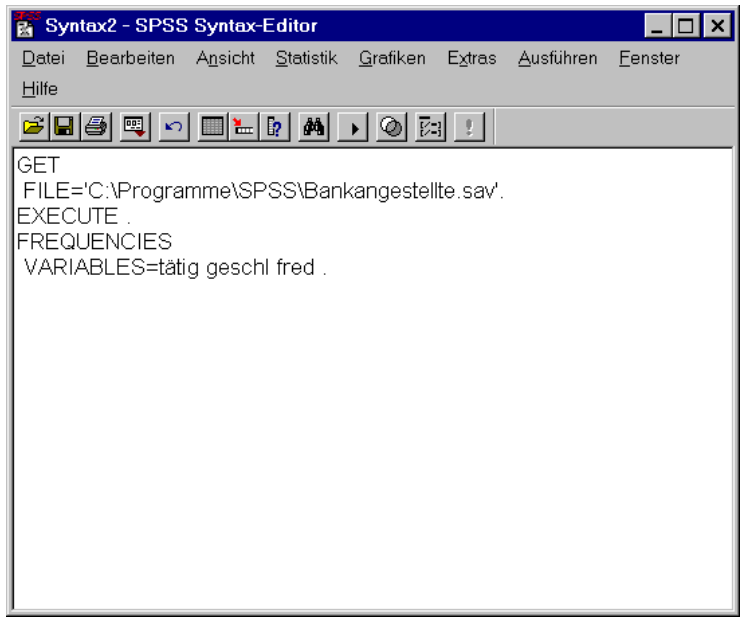
Bearbeiten der Syntax in einer Journaldatei

In der Standardeinstellung werden alle in einer Sitzung ausgeführten Befehle in einer Journaldatei mit dem Namen *spss.jnl* aufgezeichnet. (Dies wird mit dem Befehl "Optionen" aus dem Menü "Bearbeiten" eingestellt). Sie können die Journaldatei bearbeiten und als Syntaxdatei speichern. So können Sie eine vorher ausgeführte Analyse wiederholen oder die Analyse als automatisierten Job mit dem Produktionsmodus ausführen.

Die Journaldatei ist eine Textdatei, die genau wie jede andere Datei bearbeitet werden kann. Da neben der Befehlssyntax auch Fehlermeldungen und Warnungen in der Journaldatei aufgezeichnet werden, müssen Sie alle Fehlermeldungen und Warnungen entfernen, bevor Sie die Syntaxdatei speichern. Beachten sie, daß Fehler behoben werden müssen, da der Job sonst nicht erfolgreich ausgeführt wird.

Speichern Sie die bearbeitete Journaldatei unter einem anderen Namen. Die Journaldatei wird in jeder Sitzung automatisch ergänzt oder überschrieben. Falls Sie versuchen, denselben Dateinamen für eine Syntaxdatei und eine Journaldatei zu verwenden, können sich Probleme ergeben.

Abbildung 13-3
Bearbeiten der Journaldatei



So bearbeiten Sie die Syntax in einer Journaldatei:

- Wählen Sie zum Öffnen der Journaldatei die folgenden Befehle aus den Menüs aus:

  Datei
  Öffnen
  Andere...

- Suchen und öffnen Sie die Journaldatei (*spss.jnl* befindet sich in der Standardeinstellung im Verzeichnis *temp*).

  Wählen Sie im Feld “Dateityp” die Option Alle Dateien (*.*), oder geben Sie im Feld “Dateiname” den Begriff *.jnl ein, damit in der Dateiliste Journaldateien angezeigt

werden. Falls Sie die Datei nicht finden, verwenden Sie “Optionen” aus dem Menü “Bearbeiten”, um zu überprüfen, wo das Journal in Ihrem System gespeichert wurde.

- Bearbeiten Sie die Datei, um Fehlermeldungen oder Warnungen zu entfernen, die mit dem Zeichen “>” angezeigt werden.

- Speichern Sie die Datei unter einem anderen Namen. (Empfehlenswert ist ein Dateiname mit der Erweiterung *.sps*, der Standarderweiterung für Syntaxdateien.)

So führen Sie Befehlssyntax aus

- Markieren Sie die Befehle, die Sie im Syntax-Fenster ausführen möchten.

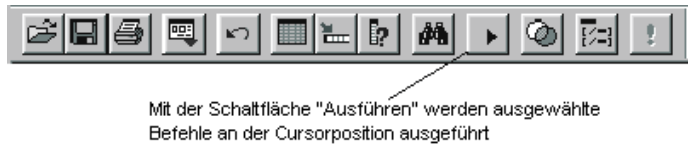
- Klicken Sie auf das Schaltfeld “Ausführen” (das nach rechts zeigende Dreieck) auf der Symbolleiste des Syntax-Editors.

  oder

- Wählen Sie einen der Befehle aus dem Menü “Ausführen”.
  - **Alles.** Führt alle Befehle im Syntax-Fenster aus.
  - **Auswahl.** Führt die aktuell ausgewählten Befehle aus. Dies umfaßt alle auch nur teilweise markierten Befehle.
  - **Aktuell.** Führt den Befehl aus, bei dem sich der Cursor gerade befindet.
  - **Bis Ende.** Führt alle Befehle von der aktuellen Cursorposition bis zum Ende der Befehlssyntax aus.

  Das Schaltfeld “Ausführen” auf der Symbolleiste des Syntaxeditors führt die ausgewählten Befehle aus. Falls Sie keine Auswahl getroffen haben, wird der Befehl, bei dem sich der Cursor gerade befindet, ausgeführt.

Abbildung 13-4
Symbolleiste des Syntax-Editors



Mehrere Execute-Befehle

Syntax, die aus Dialogfeldern eingefügt bzw. aus dem Protokoll oder Journal kopiert wird, kann Befehle EXECUTE enthalten. Wenn Sie mehrere Befehle von einem Syntax-Fenster aus ausführen, ist nur ein Befehl EXECUTE notwendig. Mehrere solcher Befehle setzen die Leistung herab, da bei diesem Befehl die gesamte Datendatei gelesen wird.

- Wenn der letzte Befehl in der Syntaxdatei ein Befehl ist, bei dem die Datendatei gelesen wird, etwa eine Statistik- oder Grafik-Prozedur, sind keine EXECUTE-Befehle erforderlich. Diese können daher gelöscht werden.
- Falls Sie nicht sicher sind, ob die Datendatei beim letzten Befehl gelesen wird, können Sie in den meisten Fällen alle EXECUTE-Befehle, mit Ausnahme des letzten, in der Syntaxdatei löschen.

Intervallfunktionen

Eine wichtige Ausnahme stellen Transformationsbefehle dar, die Intervallfunktionen beinhalten. Bei einer Reihe von Transformationsbefehlen ohne trennenden Befehl EXECUTE oder andere Befehle, die die Daten auslesen, werden die Intervallfunktionen unabhängig von der Befehlsreihenfolge nach allen anderen Transformationen berechnet. Ein Beispiel:

```
COMPUTE lagvar=LAG(var1)
COMPUTE var1=var1*2
```

und

```
COMPUTE lagvar=LAG(var1)
EXECUTE
COMPUTE var1=var1*2
```

führen zu höchst unterschiedlichen Ergebnissen für den Wert von *lagvar*, da bei ersterem der transformiert Wert von *var1* verwendet wird und bei letzterem der ursprüngliche Wert.

Kapitel
14

Häufigkeiten

Die Prozedur “Häufigkeiten” stellt Statistiken und grafische Darstellungen für die Beschreibung vieler Variablentypen zur Verfügung. Die Prozedur “Häufigkeiten” ist ein guter Ausgangspunkt für die Betrachtung Ihrer Daten.

Bei Häufigkeitsberichten und Balkendiagrammen können Sie die unterschiedlichen Werte in aufsteigender oder absteigender Reihenfolge anordnen oder die Kategorien nach deren Häufigkeiten ordnen. Der Häufigkeitsbericht kann unterdrückt werden, wenn für eine Variable viele unterschiedliche Werte vorhanden sind. Sie können Diagramme mit Häufigkeiten (die Standardeinstellung) oder Prozentsätzen beschriften.

Beispiel. Wie sind die Kunden eines Unternehmens nach Industriezweigen verteilt? Sie können aus Ihren Ausgabedaten ersehen, daß 37,5 % Ihrer Kunden zu staatlichen Behörden gehören, 24,9 % zu Unternehmen der freien Wirtschaft, 28,1 % zu akademischen Institutionen und 9,4 % zum Gesundheitswesen. Bei stetigen quantitativen Daten wie Verkaufserlösen könnten Sie beispielsweise ersehen, dass sich der durchschnittliche Produktverkauf auf $3.576 bei einer Standardabweichung von $1.078 beläuft.

Statistiken und Diagramme. Häufigkeiten, Prozentsätze, kumulierte Prozentsätze, Mittelwert, Median, Modalwert, Summe, Standardabweichung, Varianz, Spannweite, Minimum und Maximum, Standardfehler des Mittelwerts, Schiefe und Kurtosis (beide mit Standardfehler), Quartile, benutzerdefinierte Perzentile, Balkendiagramme, Kreisdiagramme und Histogramme.

Daten. Verwenden Sie zum Kodieren kategorialer Variablen numerische Codes oder kurze Strings (nominales oder ordinales Niveau der Werte).

Annahmen. Die Tabellen und Prozentsätze stellen nützliche Beschreibungen für Daten aus allen Verteilungen zur Verfügung, insbesondere für Variablen mit geordneten oder ungeordneten Kategorien. Die meisten der optionalen Auswertungsstatistiken, wie zum Beispiel der Mittelwert und die Standardabweichung, gehen von der

Normalverteilung aus und können auf quantitative Variablen mit symmetrischen Verteilungen angewendet werden. Robuste Statistiken, wie zum Beispiel Median, Quartile und Perzentile, sind für quantitative Variablen geeignet, die nur möglicherweise die Annahme erfüllen, daß eine Normalverteilung gilt.

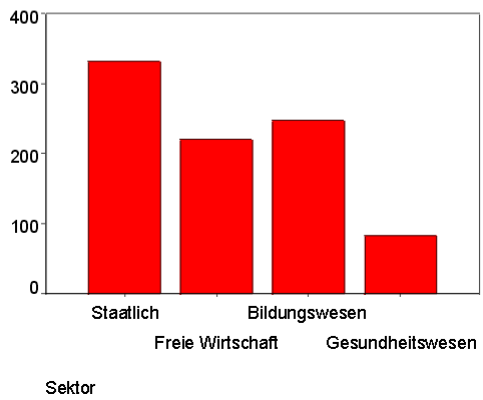
Abbildung 14-1
Ausgabe von Häufigkeiten

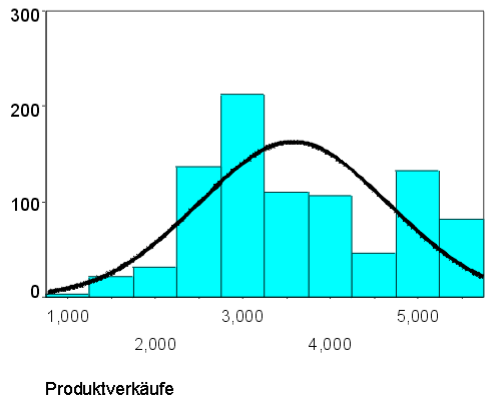
Sektor

	Häufigkeiten	Prozent	Gültige Prozent	Kumulative Prozent
Staatlich	331	37.5	37.5	37.5
Freie Wirtschaft	220	24.9	24.9	62.5
Bildungswesen	248	28.1	28.1	90.6
Gesundheitswesen	83	9.4	9.4	100.0
Gesamt	882	100.0	100.0	

Statistiken

	Mittelwert	Median	Standard-abweichung
Produkt-verkäufe	$3,576.52	$3,417.50	$1,077.84



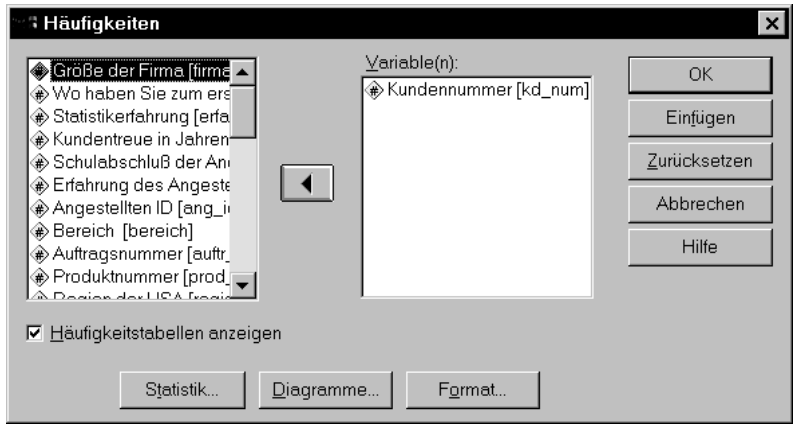


So erstellen Sie Häufigkeitstabellen

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Deskriptive Statistiken
    Häufigkeiten...

Abbildung 14-2
Dialogfeld "Häufigkeiten"



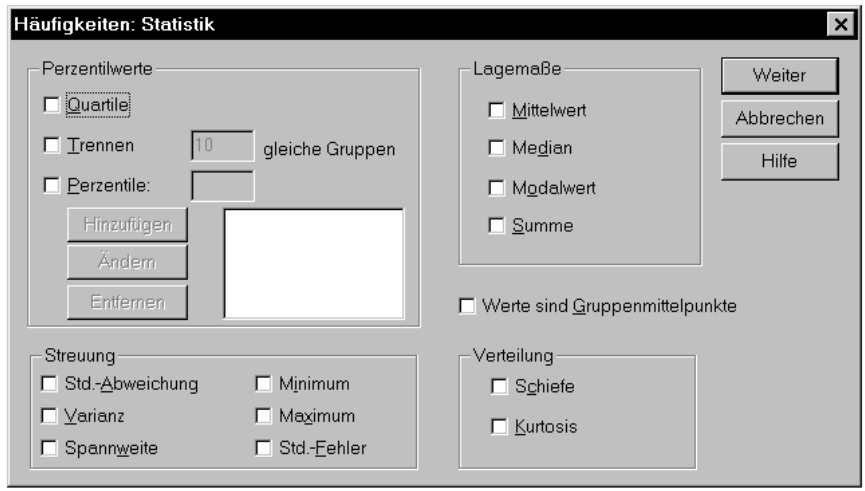
▶ Wählen Sie mindestens eine kategoriale oder quantitative Variable aus.

Die folgenden Optionen sind verfügbar:

- Deskriptive Statistiken für quantitative Variablen erhalten Sie, indem Sie auf Statistik klicken.
- Balkendiagramme, Kreisdiagramme oder Histogramme erhalten Sie, indem Sie auf Diagramme klicken.
- Sie können die Reihenfolge der angezeigten Ergebnisse ändern, indem Sie auf Format klicken.

Häufigkeiten: Statistik

Abbildung 14-3
Dialogfeld "Häufigkeiten: Statistik"



Perzentilwerte. Dies sind Werte einer quantitativen Variablen, welche die geordneten Daten in Gruppen unterteilen, so daß ein bestimmter Prozentsatz darüber und ein bestimmter Prozentsatz darunter liegt. Quartile (die 25., 50. und 75. Perzentile) unterteilen die Beobachtung in vier gleich große Gruppen. Falls Sie eine gleiche Anzahl von Gruppen wünschen, die von vier abweicht, klicken Sie auf Trennen, und geben Sie eine Anzahl für "gleiche Gruppen" ein. Sie können auch individuelle Perzentile festlegen (zum Beispiel das 95. Perzentil, also der Wert, unter dem 95 % der Beobachtungen liegen).

Lagemaße. Statistiken, welche die Lage der Verteilung beschreiben, sind Mittelwert, Median, Modalwert und Summe aller Werte.

- **Mittelwert.** Ein Lagemaß. Das arithmetische Mittel, d. h. die Summe geteilt durch die Anzahl der Fälle.
- **Median.** Der Wert, über und unter dem jeweils die Hälfte aller Fälle liegt; das 50. Perzentil. Bei einer geraden Anzahl von Fällen ist der Median der Mittelwert der zwei mittleren Fälle, wenn diese auf- oder absteigend sortiert sind. Der Median ist ein gegenüber Ausreißern unempfindliches Lagemaß - anders als der Mittelwert, der durch wenige extrem hohe oder niedrige Werte beeinflußt wird.

- **Modalwert.** Der am häufigsten auftretende Wert. Wenn mehrere Werte die gleiche größte Häufigkeit aufweisen, stellen beide Werte Modalwerte dar. Die Prozedur "Häufigkeiten" meldet bei mehreren Modalwert nur den kleinsten.
- **Summe.** Die Summe der Werte über alle Fälle mit nichtfehlenden Werten.

Streuung. Statistiken, welche die Menge an Variation oder die Streubreite in den Daten messen, sind Standardabweichung, Varianz, Spannweite, Minimum, Maximum und Standardfehler des Mittelwerts.

- **Std.abweichung.** Ein Maß für die Streuung um den Mittelwert. Bei einer Normalverteilung liegen 68% der Fälle im Bereich von einer Standardabweichung um den Mittelwert und 95% der Fälle im Bereich von zwei Standardabweichungen. Wenn z. B. für das Alter der Mittelwert 45 und die Standardabweichung 10 beträgt, würden bei Normalverteilung 95% der Fälle zwischen 25 und 65 liegen.
- **Varianz.** Ein Maß der Streuung um den Mittelwert. Es ist gleich dem Quotienten aus der Summe der quadrierten Abweichung vom Mittelwert und der um 1 verringerten Fallanzahl. Die Varianz wird in Einheiten gemessen, die das Quadrat der Einheiten der Variablen selbst sind.
- **Spannweite.** Die Differenz zwischen den größten und kleinsten Werten einer numerischen Variablen; Maximalwert minus Minimalwert.
- **Minimum.** Der kleinste Wert einer numerischen Variablen.
- **Maximum.** Der größte Wert einer numerischen Variablen.
- **Standardfehler des Mittelwerts.** Ein Maß für die mögliche Variation des Mittelwertes zwischen aus derselben Verteilung stammenden Stichproben. Er kann zum groben Vergleich des beobachteten Mittelwertes mit einem hypothetischen Wert verwendet werden (d. h. es kann geschlossen werden, daß die Werte verschieden sind, wenn das Verhältnis der Differenz zum Standardfehler geringer ist als -2 oder größer als +2).

Verteilung. Schiefe und Kurtosis sind Statistiken, die Form und Symmetrie der Verteilung beschreiben. Diese Statistiken werden mit ihren Standardfehlern angezeigt.

- **Schiefe.** Ein Maß für die Asymmetrie einer Verteilung. Die Normalverteilung ist symmetrisch, ihre Schiefe hat den Wert Null. Eine Verteilung mit einer deutlichen positiven Schiefe läuft nach rechts lang aus (langer rechter Schwanz), eine Verteilung mit einer deutlichen negativen Schiefe läuft nach links lang aus (langer linker Schwanz). Als Faustregel kann man verwenden, dass ein

Schiefe-Wert, der mehr als doppelt so groß ist wie sein Standardfehler für eine Abweichung von der Symmetrie spricht.

- **Kurtosis.** Ein Maß dafür, wie sich die Beobachtungen um einen zentralen Punkt gruppieren. Bei einer Normalverteilung ist der Wert der Kurtosis gleich 0. Bei positiver Kurtosis gruppieren sich die Beobachtungen dichter als bei der Normalverteilung und haben längere Flügel. Bei negativer Kurtosis gruppieren sich die Beobachtungen weniger dicht zusammen und haben kürzere Flügel.

Werte sind Gruppenmittelpunkte. Falls die Werte in den Daten Gruppenmittelpunkte sind (wenn zum Beispiel das Alter aller Personen in den Dreißigern mit dem Wert 35 kodiert ist), wählen Sie diese Option, um den Median und das Perzentil für die ursprünglichen, nicht gruppierten Daten berechnen zu lassen.

Häufigkeiten: Diagramme

Abbildung 14-4
Dialogfeld "Häufigkeiten: Diagramme"

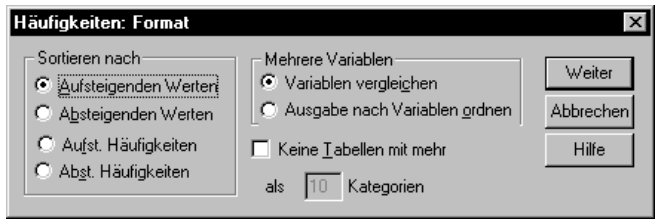
Diagrammtyp. In einem Kreisdiagramm wird der Anteil der Teile an einem Ganzen angezeigt. Jedes Segment eines Kreisdiagramms entspricht einer durch eine einzelne Gruppenvariable definierten Gruppe. In einem Balkendiagramm wird die Anzahl für jeden unterschiedlichen Wert oder jede unterschiedliche Kategorie als separater Balken angezeigt, wodurch Sie Kategorien visuell vergleichen können. Auch Histogramme enthalten Balken, diese sind jedoch an einer Skala mit gleichen Abständen ausgerichtet. Die Höhe jedes Balkens gibt die Anzahl der Werte einer quantitativen Variablen wieder, die innerhalb des Intervalls liegen. In einem Histogramm werden Form, Mittelpunkt und die Streubreite der Verteilung

angezeigt. Eine über das Histogramm gelegte Normalverteilungskurve erleichtert die Beurteilung, ob die Daten normalverteilt sind.

Diagrammwerte. Bei Balkendiagrammen kann die Skalenachse mit Häufigkeiten oder Prozentwerten beschriftet werden.

Häufigkeiten: Format

Abbildung 14-5
Dialogfeld "Häufigkeiten: Format"



Sortieren nach. Die Häufigkeitstabelle kann entsprechend der tatsächlichen Werte der Daten oder entsprechend der Anzahl (Häufigkeit des Vorkommens) dieser Werte geordnet werden. Die Tabelle kann entweder in aufsteigender oder in absteigender Reihenfolge angeordnet werden. Wenn Sie allerdings ein Histogramm oder Perzentile anfordern, wird in der Prozedur "Häufigkeiten" davon ausgegangen, daß die Variable quantitativ ist. Die Werte werden dann in aufsteigender Reihenfolge angezeigt.

Mehrere Variablen. Wenn Sie Statistiktabellen für multiple Variablen erzeugen, können Sie entweder alle Variablen in einer einzigen Tabelle (Variablen vergleichen) oder eine eigene Statistiktabelle für jede Variable (Ausgabe nach Variablen ordnen) anzeigen.

Keine Tabellen mit mehr als n Kategorien. Diese Option verhindert die Anzeige von Tabellen mit mehr als der angegebenen Anzahl von Werten.

Kapitel
15

Deskriptive Statistiken

Mit der Prozedur "Deskriptive Statistiken" werden in einer einzelnen Tabelle univariate Auswertungsstatistiken für verschiedene Variablen angezeigt und standardisierte Werte (*Z*-Werte) errechnet. Variablen können folgendermaßen geordnet werden: nach der Größe ihres Mittelwerts (in aufsteigender oder absteigender Reihenfolge), alphabetisch oder in der Reihenfolge, in der sie ausgewählt wurden (dies ist die Standardeinstellung).

Wenn *Z*-Werte gespeichert werden, werden sie zu den Daten im Daten-Editor hinzugefügt und stehen dann für SPSS-Diagramme, Auflistungen von Daten und Analysen zur Verfügung. Wenn Variablen in verschiedenen Einheiten aufgezeichnet werden (zum Beispiel Bruttoinlandsprodukt pro Kopf der Bevölkerung und Prozentsatz der Alphabetisierung), werden die Variablen durch eine *Z*-Wert-Transformation zur Erleichterung des visuellen Vergleichs auf einer gemeinsamen Skala angeordnet.

Beispiel. Sie zeichnen über mehrere Monate den täglichen Umsatz jedes einzelnen Angestellten der Verkaufsabteilung auf (z. B. ein Eintrag für Herbert, ein Eintrag für Sabine, ein Eintrag für Joachim usw.), sodass jeder Fall in Ihren Daten den täglichen Umsatz jedes Angestellten enthält. Mit der Prozedur "Deskriptive Statistik" wird für Sie jetzt der durchschnittliche Tagesumsatz der einzelnen Angestellten berechnet und das Ergebnis vom höchsten durchschnittlichen Umsatz zum niedrigsten durchschnittlichen Umsatz geordnet.

Statistiken. Stichprobengröße, Mittelwert, Minimum, Maximum, Standardabweichung, Varianz, Spannweite, Summe, Standardfehler des Mittelwerts und Kurtosis und Schiefe mit den Standardfehlern.

Daten. Verwenden Sie numerische Variablen, nachdem Sie diese im Diagramm auf Aufzeichnungsfehler, Ausreißer und Unregelmäßigkeiten in der Verteilung untersucht haben. Die Prozedur "Deskriptive Statistiken" ist für große Dateien (mit Tausenden von Fällen) besonders effektiv.

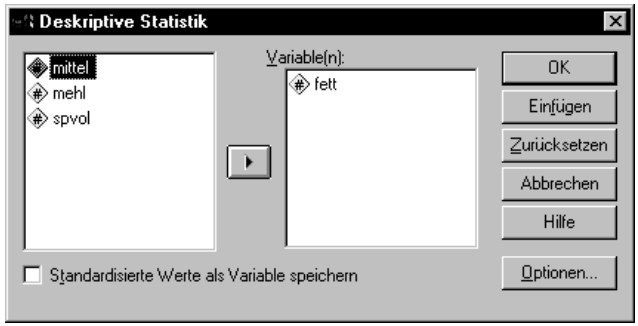
Annahmen. Die meisten verfügbaren Statistiken (einschließlich *Z*-Werte) basieren auf der Annahme, dass die Daten normalverteilt sind, und sind für quantitative Variablen (mit Intervall- oder Verhältnis-Meßniveau) mit symmetrischen Verteilungen geeignet. Vermeiden Sie Variablen mit ungeordneten Kategorien oder schiefen Verteilungen. Die Verteilung der *Z*-Werte hat dieselbe Form wie die ursprünglichen Daten; daher bietet das Berechnen von *Z*-Werten keine Abhilfe bei problematischen Daten.

So lassen Sie deskriptive Statistiken berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Deskriptive Statistiken
    Deskriptive Statistiken...

Abbildung 15-1
Dialogfeld "Deskriptive Statistik"



▶ Wählen Sie mindestens eine Variable aus.

Die folgenden Optionen sind verfügbar:

- Wählen Sie Standardisierte Werte als Variable speichern, um *Z*-Werte als neue Variablen zu speichern.
- Optionale Statistiken und die Reihenfolge der Anzeige steuern Sie, indem Sie auf Optionen klicken.

Deskriptive Statistik: Optionen

Abbildung 15-2
Dialogfeld "Deskriptive Statistik: Optionen"

Mittelwert und Summe. In der Standardeinstellung wird der Mittelwert bzw. das arithmetische Mittel angezeigt.

Streuung. Zu den Statistiken, welche die Streubreite oder die Variation in den Daten messen, gehören Standardabweichung, Varianz, Spannweite, Minimum, Maximum und Standardfehler des Mittelwerts.

- **Std.abweichung.** Ein Maß für die Streuung um den Mittelwert. Bei einer Normalverteilung liegen 68% der Fälle im Bereich von einer Standardabweichung um den Mittelwert und 95% der Fälle im Bereich von zwei Standardabweichungen. Wenn z. B. für das Alter der Mittelwert 45 und die Standardabweichung 10 beträgt, würden bei Normalverteilung 95% der Fälle zwischen 25 und 65 liegen.
- **Varianz.** Ein Maß der Streuung um den Mittelwert. Es ist gleich dem Quotienten aus der Summe der quadrierten Abweichung vom Mittelwert und der um 1 verringerten Fallanzahl. Die Varianz wird in Einheiten gemessen, die das Quadrat der Einheiten der Variablen selbst sind.
- **Spannweite.** Die Differenz zwischen den größten und kleinsten Werten einer numerischen Variablen; Maximalwert minus Minimalwert.
- **Minimum.** Der kleinste Wert einer numerischen Variablen.

- **Maximum.** Der größte Wert einer numerischen Variablen.
- **Standardfehler des Mittelwerts.** Ein Maß für die mögliche Variation des Mittelwertes zwischen aus derselben Verteilung stammenden Stichproben. Er kann zum groben Vergleich des beobachteten Mittelwertes mit einem hypothetischen Wert verwendet werden (d. h. es kann geschlossen werden, daß die Werte verschieden sind, wenn das Verhältnis der Differenz zum Standardfehler geringer ist als -2 oder größer als +2).

Verteilung. Kurtosis und Schiefe sind Statistiken, die Form und Symmetrie der Verteilung charakterisieren. Diese Statistiken werden mit ihren Standardfehlern angezeigt.

- **Kurtosis.** Ein Maß dafür, wie sich die Beobachtungen um einen zentralen Punkt gruppieren. Bei einer Normalverteilung ist der Wert der Kurtosis gleich 0. Bei positiver Kurtosis gruppieren sich die Beobachtungen dichter als bei der Normalverteilung und haben längere Flügel. Bei negativer Kurtosis gruppieren sich die Beobachtungen weniger dicht zusammen und haben kürzere Flügel.
- **Schiefe.** Ein Maß für die Asymmetrie einer Verteilung. Die Normalverteilung ist symmetrisch, ihre Schiefe hat den Wert Null. Eine Verteilung mit einer deutlichen positiven Schiefe läuft nach rechts lang aus (langer rechter Schwanz), eine Verteilung mit einer deutlichen negativen Schiefe läuft nach links lang aus (langer linker Schwanz). Als Faustregel kann man verwenden, dass ein Schiefe-Wert, der mehr als doppelt so groß ist wie sein Standardfehler für eine Abweichung von der Symmetrie spricht.

Anzeigereihenfolge. In der Standardeinstellung werden die Variablen in der Reihenfolge angezeigt, in der sie ausgewählt wurden. Sie können Variablen bei Bedarf in alphabetischer Reihenfolge mit aufsteigend oder absteigend geordneten Mittelwerten anzeigen lassen.

Zusätzliche Funktionen beim Befehl DESCRIPTIVES

Mit der SPSS-Befehlssprache verfügen Sie über die folgenden zusätzlichen Möglichkeiten:

- Sie können die standardisierten Werte (*Z*-Werte) selektiv für einige Variablen speichern (mit dem Unterbefehl `VARIABLES`).

- Sie können Namen für die neuen Variablen angeben, die die standardisierte Werte enthalten (mit dem Unterbefehl `VARIABLES`).
- Sie können Fälle mit fehlenden Werten in einer beliebigen Variablen aus der Analyse ausschließen (mit dem Unterbefehl `MISSING`).
- Sie können die Variablen in der Anzeige nach dem Wert einer beliebigen Statistik, nicht nur nach dem Mittelwert sortieren (mit dem Unterbefehl `SORT`).

Vollständige Informationen zur Syntax finden Sie in der *SPSS Command Syntax Reference*.

Kapitel
16

Explorative Datenanalyse

Mit der Prozedur “Explorative Datenanalyse” werden Auswertungsstatistiken und grafische Darstellungen für alle Fälle oder für separate Fallgruppen erzeugt. Es kann viele Gründe für die Verwendung der Prozedur “Explorative Datenanalyse” geben: Sichten von Daten, Erkennen von Ausreißern, Beschreibung, Überprüfung der Annahmen und Charakterisieren der Unterschiede zwischen Teilgrundgesamtheiten (Fallgruppen). Beim Sichten der Daten können Sie ungewöhnliche Werte, Extremwerte, Lücken in den Daten oder andere Auffälligkeiten erkennen. Durch die explorative Datenanalyse können Sie sich vergewissern, ob die für die Datenanalyse vorgesehenen statistischen Methoden geeignet sind. Die Untersuchung kann ergeben, daß Sie die Daten transformieren müssen, falls die Methode eine Normalverteilung erfordert. Sie können sich statt dessen auch für die Verwendung nichtparametrischer Tests entscheiden.

Beispiel. Betrachten Sie die Verteilung der Lernzeiten für Ratten im Labyrinth mit vier verschiedenen Schwierigkeitsgraden. Zu jeder der vier Gruppen können Sie ablesen, ob die Zeiten annähernd normalverteilt und die vier Varianzen gleich sind. Sie können auch die Fälle mit den fünf längsten und den fünf kürzesten Zeiten bestimmen. Sie können die Verteilung der Lernzeiten für jede Gruppe mit Boxplots und Stengel-Blatt-Diagrammen grafisch auswerten.

Statistiken und Diagramme. Mittelwert, Median, 5% getrimmtes Mittel, Standardfehler, Varianz, Standardabweichung, Minimum, Maximum, Spannweite, interquartiler Bereich, Schiefe und Kurtosis und deren Standardfehler, Konfidenzintervall für den Mittelwert (und angegebenes Konfidenzniveau), Perzentile, M-Schätzer nach Huber, Andrew-Wellen-Schätzer, M-Schätzer nach Hampel, Tukey-Biweight-Schätzer, die fünf größten und die fünf kleinsten Werte, die Kolmogorov-Smirnov-Statistik mit Lilliefors-Signifikanzniveau zum Prüfen der Normalverteilung und die Shapiro-Wilk-Statistik. Boxplots, Stengel-Blatt-Diagramme, Histogramme, Normalverteilungsdiagramme und Diagramme der Streubreite gegen das mittlere Niveau mit Levene-Test und Transformationen.

Daten. Die Prozedur "Explorative Datenanalyse" kann für quantitative Variablen (mit Intervall- oder Verhältnis-Meßniveau) verwendet werden. Eine Faktorvariable (zum Aufteilen der Daten in Fallgruppen) muß eine sinnvolle Anzahl von unterschiedlichen Werten (Kategorien) enthalten. Diese Werte können kurze Strings oder numerische Werte sein. Die Fallbeschriftungsvariable, die für die Beschriftung von Ausreißern in Boxplots verwendet wird, kann ein kurzer String, ein langer String (die ersten 15 Zeichen) oder numerisch sein.

Annahmen. Ihre Daten müssen nicht symmetrisch oder normalverteilt sein.

Abbildung 16-1
Ausgabe der explorativen Datenanalyse

Univariate Statistiken

			Zeit			
			Plan			
			1	2	3	4
Mittelwert		Statistik	2,760	4,850	6,900	9,010
		Standardf ehler	,165	,422	,445	,289
95% Konfidenzintervall des Mittelwerts	Untergrenze	Statistik	2,387	3,895	5,893	8,357
	Obergrenze	Statistik	3,133	5,805	7,907	9,663
5% getrimmtes Mittel		Statistik	2,761	4,889	6,911	8,994
Median		Statistik	2,850	4,900	7,050	9,000
Varianz		Statistik	,272	1,783	1,982	,834
Standardabweichung		Statistik	,521	1,335	1,408	,913
Minimum		Statistik	2,0	2,3	4,5	7,8
Maximum		Statistik	3,5	6,7	9,1	10,5
Spannweite		Statistik	1,5	4,4	4,6	2,7
Interquartilbereich		Statistik	,925	2,250	2,400	1,650
Schiefe		Statistik	-,116	-,559	-,197	,219
		Standardf ehler	,687	,687	,687	,687
Kurtosis		Statistik	-1,210	-,104	-,606	-1,350
		Standardf ehler	1,334	1,334	1,334	1,334

Extremwerte

			Fallnummer	Plan	Wert
Zeit	Größte Werte	1	31	4	10,5
		2	33	4	9,9
		3	39	4	9,8
		4	32	4	9,5
		5	36	4	9,3
	Kleinste Werte	1	2	1	2,0
		2	7	1	2,1
		3	1	1	2,3
		4	11	2	2,3
		5	3	1	2,5

Stengel-Blatt (Stem & Leaf)

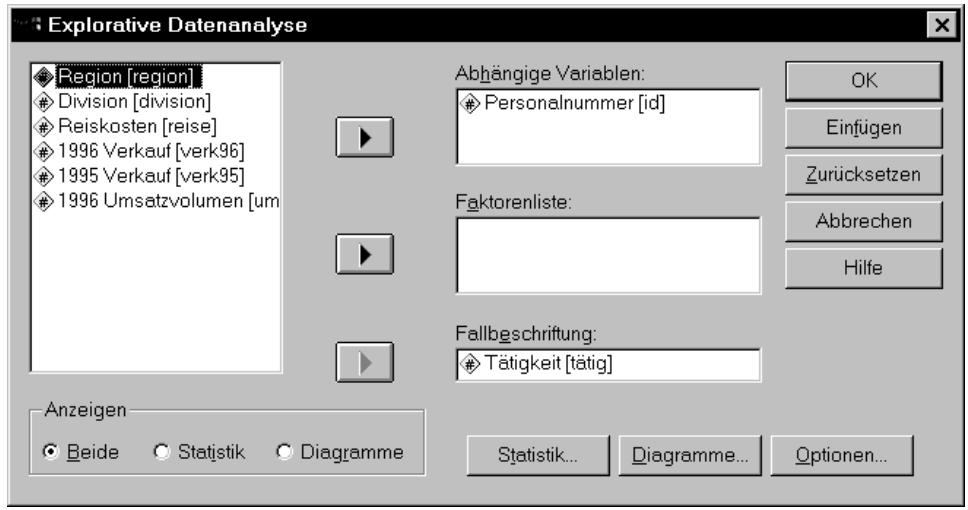
```
  7,00      2 , 0133589
  6,00      3 , 014577
  3,00      4 , 568
  5,00      5 , 05779
  4,00      6 , 1379
  3,00      7 , 268
  6,00      8 , 012237
  5,00      9 , 13589
  1,00     10 , 5

Breite des Stengels:     1,0
Jedes Blatt:             1 Fall/Fälle
```

So führen Sie eine explorative Datenanalyse aus:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Deskriptive Statistiken
    Explorative Datenanalyse

Abbildung 16-2
Dialogfeld "Explorative Datenanalyse"



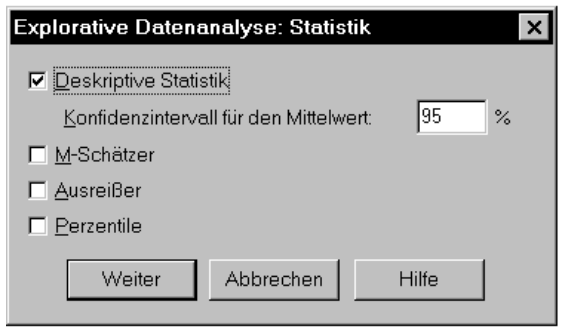
▶ Wählen Sie eine oder mehrere abhängige Variablen aus.

Die folgenden Optionen sind verfügbar:

- Auswählen einer oder mehrerer Faktorvariablen, mit deren Werten Fallgruppen definiert werden.
- Auswählen einer Identifizierungsvariablen für die Beschriftung von Fällen.
- Zugriff auf robuste Schätzer, Ausreißer, Perzentile und Häufigkeitstabellen erhalten Sie, indem Sie auf Statistik klicken.
- Zugriff auf Histogramme, Normalverteilungsdiagramme und Tests sowie Diagramme der Streubreite gegen das mittlere Niveau mit Levene-Statistik erhalten Sie, indem Sie auf Diagramme klicken.
- Sie können die Behandlung fehlender Werte festlegen, indem Sie auf Optionen klicken.

Explorative Datenanalyse: Statistik

Abbildung 16-3
Dialogfeld "Explorative Datenanalyse: Statistik"



Deskriptive Statistiken. In der Standardeinstellung werden Lage- und Streuungsmaße angezeigt. Mit den Lagemaßen wird die Lage der Verteilung angegeben. Dazu gehören Mittelwert, Median und 5% getrimmtes Mittel. Mit den Maßen für Streuung werden Unähnlichkeiten der Werte angezeigt. Diese umfassen Standardfehler, Varianz, Standardabweichung, Minimum, Maximum, Spannweite und den Interquartilbereich. Die beschreibenden Statistiken enthalten auch Maße der Verteilungsform. Schiefe und Kurtosis werden mit den jeweiligen Standardfehlern angezeigt. Das 95%-Konfidenzintervall für den Mittelwert wird ebenfalls angezeigt. Sie können auch ein anderes Konfidenzniveau angeben.

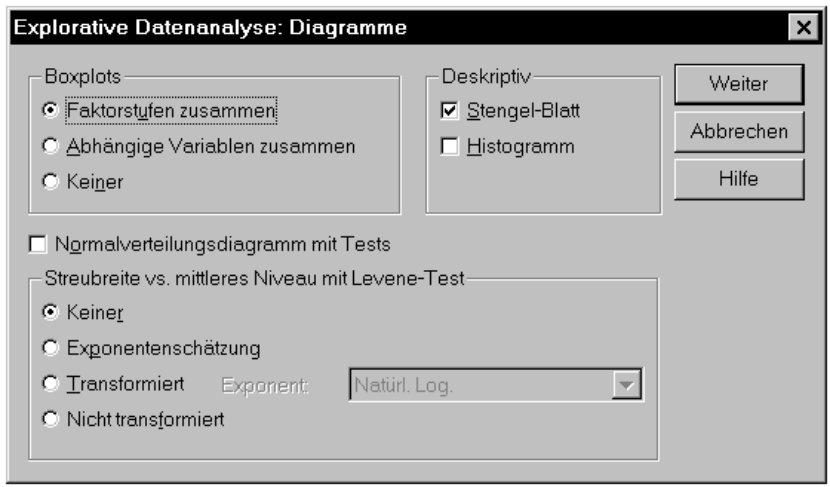
M-Schätzer. Robuste Alternativen zu Mittelwert und Median der Stichprobe zum Schätzen der Lage. Die berechneten Schätzer unterscheiden sich in den Gewichtungen, die sie den Fällen zuweisen. M-Schätzer nach Huber, Andrew-Wellen-Schätzer, M-Schätzer nach Hampel und Tukey-Biweight-Schätzer werden angezeigt.

Ausreißer. Hier werden die fünf größten und die fünf kleinsten Werte mit Fallbeschriftungen angezeigt.

Perzentile. Hier werden die Werte für die 5., 10., 25., 50., 75., 90. und 95. Perzentile angezeigt.

Explorative Datenanalyse: Diagramme

Abbildung 16-4
Dialogfeld "Explorative Datenanalyse: Diagramme"



Boxplots. Mit diesen Optionen legen Sie fest, wie Boxplots bei mehr als einer abhängigen Variablen angezeigt werden. Mit Faktorstufen zusammen wird eine getrennte Anzeige für jede abhängige Variable erzeugt. In einer Anzeige werden Boxplots für alle durch eine Faktorvariable definierten Gruppen angezeigt. Mit Abhängige Variablen zusammen wird für jede durch eine Faktorvariable definierte Gruppe eine getrennte Anzeige erzeugt. In einer Anzeige werden Boxplots für alle abhängigen Variablen in einer Anzeige nebeneinander dargestellt. Diese Anzeige ist insbesondere nützlich, wenn verschiedene Variablen ein einziges, zu unterschiedlichen Zeiten gemessenes Merkmal darstellen.

Deskriptive Statistik. Im Gruppenfeld "Deskriptive Statistik" können Sie Stengel-Blatt-Diagramme und Histogramme auswählen.

Normalverteilungsdiagramme mit Tests. Hier werden Normalverteilungsdiagramme und trendbereinigte Normalverteilungsdiagramme angezeigt. Die Kolmogorov-Smirnov-Statistik mit einem Signifikanzniveau nach Lilliefors für den Test auf Normalverteilung wird angezeigt. Bei Angabe von nicht-ganzzahligen Gewichtungen wird die Shapiro-Wilk-Statistik berechnet, wenn die gewichtete Stichprobengröße zwischen 3 und 50 liegt. Bei keinen oder ganzzahligen

Gewichtungen wird die Statistik berechnet, wenn die gewichtete Stichprobengröße zwischen 3 und 5000 liegt.

Streubreite vs. mittleres Niveau mit Levene-Test. Hiermit legen Sie fest, wie Daten für Diagramme der Streubreite versus mittleres Niveau transformiert werden. Für alle Diagramme der Streubreite versus mittleres Niveau werden die Steigung der Regressionsgeraden und der Levene-Test auf Homogenität der Varianz angezeigt. Wenn Sie eine Transformation auswählen, liegen dem Levene-Test die transformierten Daten zugrunde. Wenn keine Faktorvariable ausgewählt wurde, werden keine Diagramme der Streubreite versus mittleres Niveau erstellt. Mit der Exponentenschätzung wird ein Diagramm der natürlichen Logarithmen des Interquartilbereichs über die natürlichen Logarithmen des Medians für alle Zellen sowie eine Schätzung der Potenztransformation zum Erreichen gleicher Varianzen in den Zellen angefordert. Mit Diagrammen der Streubreite versus mittleres Niveau lässt sich der Exponent für Transformationen bestimmen, mit denen über Gruppen hinweg eine höhere Stabilität (höhere Gleichförmigkeit) der Varianzen erreicht wird. Mit Transformiert können Sie einen alternativen Exponenten auswählen, eventuell gemäß der Empfehlung der Exponentenschätzung, und Diagramme der transformierten Daten erzeugen. Der Interquartilbereich und der Median der transformierten Daten werden grafisch dargestellt. Mit Nicht transformiert werden Diagramme der Rohdaten erstellt. Dies entspricht einer Transformation mit einem Exponenten gleich 1.

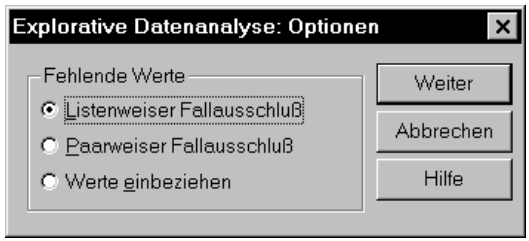
Explorative Datenanalyse: Potenztransformationen

Dies sind die Potenztransformationen für Diagramme der Streubreite versus mittleres Niveau. Für die Transformation von Daten muß ein Exponent ausgewählt werden. Sie können eine der folgenden Möglichkeiten wählen:

- **Natürlicher Logarithmus.** Transformation mit natürlichem Logarithmus. Dies ist die Standardeinstellung.
- **1/Quadratwurzel.** Zu jedem Datenwert wird der reziproke Wert der Quadratwurzel berechnet.
- **Reziprok.** Der reziproke Wert jedes Datenwerts wird berechnet.
- **Quadratwurzel.** Die Quadratwurzel jedes Datenwerts wird berechnet.
- **Quadratisch.** Jeder Datenwert wird quadriert.
- **Kubisch.** Es wird die dritte Potenz jedes Datenwerts errechnet.

Explorative Datenanalyse: Optionen

Abbildung 16-5
Dialogfeld "Explorative Datenanalyse: Optionen"



Fehlende Werte. Bestimmt die Verarbeitung fehlender Werte.

- **Listenweiser Fallausschluß.** Fälle mit fehlenden Werten für abhängige Variablen oder Faktorvariablen werden aus allen Analysen ausgeschlossen. Dies ist die Standardeinstellung.
- **Paarweiser Fallausschluß.** Fälle ohne fehlenden Werte für Variablen in einer Gruppe (Zelle) werden in die Analyse dieser Gruppe einbezogen. Der Fall kann fehlende Werte für Variablen enthalten, die in anderen Gruppen verwendet werden.
- **Werte einbeziehen.** Fehlende Werte für Faktorvariablen werden als gesonderte Kategorie behandelt. Die gesamte Ausgabe wird auch für diese zusätzliche Kategorie erstellt. Häufigkeitstabellen enthalten Kategorien für fehlende Werte. Fehlende Werte für Faktorvariablen werden aufgenommen, jedoch als fehlend beschriftet.

Zusätzliche Funktionen beim Befehl EXAMINE

In der Prozedur "Explorative Datenanalyse" wird die Befehlssyntax von `EXAMINE` verwendet. Mit der SPSS-Befehlssprache verfügen Sie über die folgenden zusätzlichen Möglichkeiten:

- Anfordern von Ausgaben und Diagrammen für Gesamtsummen neben den Ausgaben und Diagrammen für Gruppen, die durch die Faktorvariablen definiert wurden (mit dem Unterbefehl `TOTAL`).
- Angeben einer gemeinsamen Skala für eine Gruppe von Boxplots (mit dem Unterbefehl `SCALE`).

- Angeben von Interaktionen der Faktorvariablen (mit dem Unterbefehl `VARIABLES`).
- Angeben von anderen Perzentilen als in der Standardeinstellung (mit dem Unterbefehl `PERCENTILES`).
- Berechnen der Perzentile nach fünf Methoden (mit dem Unterbefehl `PERCENTILES`).
- Angeben einer Potenztransformation für Diagramme der Streubreite gegen das mittlere Niveau (mit dem Unterbefehl `PLOT`).
- Angeben der Anzahl von Extremwerten, die angezeigt werden sollen (mit dem Unterbefehl `STATISTICS`).
- Angeben der Parameter für die M-Schätzer, den robusten Schätzern der Lage (mit dem Unterbefehl `MESTIMATORS`).

Vollständige Informationen zur Syntax finden Sie in der *SPSS Command Syntax Reference*.

Kapitel
17

Kreuztabellen

Mit der Prozedur “Kreuztabellen” erzeugen Sie Zweifach- und Mehrfach-Tabellen. Es stehen eine Vielzahl von Tests und Zusammenhangsmaßen für Zweifach-Tabellen zur Verfügung. Welcher Test oder welches Maß verwendet wird, hängt von der Struktur der Tabelle ab und davon, ob die Kategorien geordnet sind.

Statistiken und Zusammenhangsmaße für Kreuztabellen werden nur für Zweifach-Tabellen berechnet. Wenn Sie eine Zeile, eine Spalte und einen Schichtfaktor (Kontroll-Variable) festlegen, wird von der Prozedur “Kreuztabelle” eine separate Ausgabe mit der entsprechenden Statistik sowie den Maßen für jeden Wert des Schichtfaktors (oder eine Kombination der Werte für zwei oder mehrere Kontroll-Variablen) angezeigt. Wenn zum Beispiel *Geschlecht* ein Schichtfaktor für eine Tabelle ist, wobei *verheiratet* (Ja, Nein) gegenüber *Leben* (ist das Leben aufregend, Routine oder langweilig) untersucht wird, werden die Ergebnisse für eine Zweifach-Tabelle für weibliche Personen getrennt von den männlichen berechnet und als aufeinanderfolgende separate Ausgaben gedruckt.

Beispiel. Wie groß ist die Wahrscheinlichkeit, daß mit den Kunden aus kleineren Unternehmen beim Verkauf von Dienstleistungen (zum Beispiel Weiterbildung und Beratung) ein größerer Gewinn erzielt wird als mit den Kunden aus größeren Unternehmen? Einer Kreuztabelle könnten Sie möglicherweise entnehmen, daß die Mehrheit der kleinen Unternehmen (mit mehr als 500 Angestellten) beim Verkauf von Dienstleistungen einen hohen Gewinn erzielt, während die meisten großen Unternehmen (mit mehr als 2,500 Angestellten) dabei nur niedrige Gewinne erzielen.

Statistiken und Zusammenhangsmaße. Pearson-Chi-Quadrat, Likelihood-Quotienten-Chi-Quadrat, Zusammenhangstest linear-mit-linear, Exakter Test nach Fisher, korrigiertes Chi-Quadrat nach Yates, Pearson-*r*, Spearman-Rho, Kontingenzkoeffizient, Phi, Cramér-*V*, symmetrische und asymmetrische Lambdas, Goodman-und-Kruskal-Tau, Unsicherheitskoeffizient, Gamma, Somer-*d*, Kendall-Tau-*b*, Kendall-Tau-*c*, Eta-Koeffizient, Cohen-Kappa,

relativer Risikoschätzer, Quotenverhältnis, McNemar-Test, Cochran- und Mantel-Haenszel-Statistik.

Daten. Um die Kategorien der Tabellenvariablen zu definieren, verwenden Sie numerische Werte oder kurze String-Variablen (acht oder weniger Buchstaben). Zum Beispiel können Sie die Daten für *Geschlecht* als 1 und 2 oder als *männlich* und *weiblich* kodieren.

Annahmen. Einige Statistiken und Maße setzen geordnete Kategorien (Ordinal-Daten) oder quantitative Werte (Intervall- oder Verhältnisdaten) voraus, wie bereits im Abschnitt über Statistiken erläutert wurde. Andere sind zulässig, wenn die Tabellenvariablen über ungeordnete Kategorien verfügen (Nominal-Daten). Für Statistiken, die auf Chi-Quadrat basieren (Phi, Cramér-*V*, Kontingenzkoeffizient), sollten die Daten durch eine Zufallsstichprobe aus einer multinomialen Verteilung bezogen werden.

Anmerkung: Bei ordinalen Variablen kann es sich um numerische Codes für Kategorien (z. B. 1 = *schwach*, 2 = *mittel*, 3 = *stark*) oder um String-Werte handeln. Die alphabetische Ordnung der String-Werte gibt dabei die Reihenfolge der Kategorien vor. Bei einer String-Variablen mit den Werten *Schwach*, *Mittel* und *Stark* werden die Kategorien beispielsweise in der Reihenfolge *Mittel*, *Schwach*, *Stark* und somit falsch angeordnet. Im allgemeinen ist die Verwendung von numerischem Code für ordinale Daten günstiger.

Abbildung 17-1
Ausgabe für "Kreuztabellen"

Produktivität Service * Mitarbeiteranzahl der Firma Kreuztabelle

Anzahl

		Mitarbeiteranzahl der Firma			
		1-500	501-2,500	> 2,500	Gesamt
Produktivität Service	niedrig	200	85	135	420
	hoch	251	106	105	462
Gesamt		451	191	240	882

Chi-Quadrat-Tests

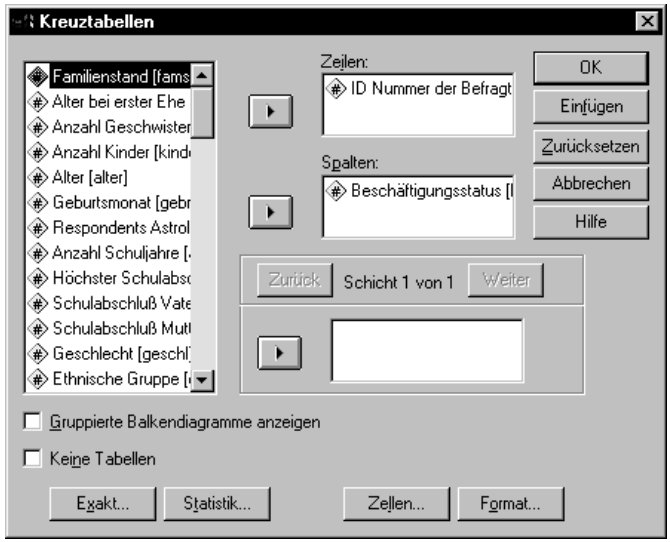
	Wert	df	Asymptotische Signifikanz (2-seitig)
Chi-Quadrat nach Pearson	9,848	2	,007
Likelihood-Quotient	9,852	2	,007
Zusammenhang linear-mit-linear	7,869	1	,005
Anzahl der gültigen Fälle	882		

So lassen Sie Kreuztabellen berechnen

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Deskriptive Statistiken
    Kreuztabellen

Abbildung 17-2
Dialogfeld "Kreuztabellen"



- Wählen Sie eine oder mehrere Zeilenvariablen und eine oder mehrere Spaltenvariablen aus.

Die folgenden Optionen sind verfügbar:

- Eine oder mehrere Kontroll-Variablen auswählen.
- Tests und Zusammenhangsmaße der Zweifach-Tabellen oder Untertabellen erhalten Sie, indem Sie auf Statistik klicken.
- Informationen zu beobachteten und erwarteten Werten, Prozentsätzen und Residuen erhalten Sie, indem Sie auf Zellen klicken.
- Durch Klicken auf Format können Sie die Reihenfolge der Kategorien festlegen.

Kreuztabellenschichten

Wenn Sie eine oder mehrere Schichtvariablen auswählen, wird für jede Kategorie jeder Schichtvariablen (Kontroll-Variablen) jeweils eine Kreuztabelle erzeugt. Wenn Sie zum Beispiel über eine Zeilenvariable, eine Spaltenvariable und eine Schichtvariable mit zwei Kategorien verfügen, erhalten Sie eine Zweifach-Tabelle für jede Kategorie der Schichtvariablen. Um eine weitere Schicht von Kontroll-Variablen

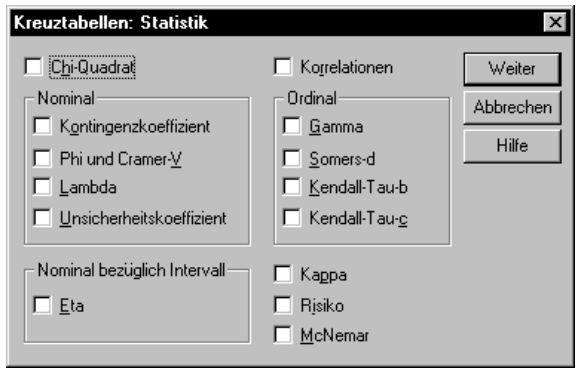
anzulegen, klicken Sie auf Weiter. Untertabellen werden für jede Kombination von Kategorien erzeugt, d. h. für jede Variable der ersten Schicht mit jeder Variablen der zweiten Schicht und so weiter. Wenn Statistiken und Zusammenhangsmaße angefordert werden, treffen diese nur auf Zweifach-Untertabellen zu.

Kreuztabellen: Gruppierte Balkendiagramme

Gruppierte Balkendiagramme anzeigen. Mit einem gruppierten Balkendiagramm können Sie Ihre Daten leichter nach Gruppen von Fällen auswerten. Für jeden Wert der Variablen, der von Ihnen unter Zeilen festgelegt wurde, wird eine Gruppe von Balken erzeugt. Die Balken in jedem Cluster werden durch die unter Spalten angegebene Variable definiert. Für jeden Wert dieser Variablen steht Ihnen ein Set unterschiedlich farbiger oder gemusterter Balken zur Verfügung. Wenn Sie unter Zeilen oder Spalten mehr als eine Variable angeben, wird für jede Kombination von zwei Variablen ein gruppiertes Balkendiagramm erzeugt.

Kreuztabellen: Statistik

Abbildung 17-3
Dialogfeld "Kreuztabellen: Statistik"



Chi-Quadrat. Für Tabellen mit zwei Zeilen und zwei Spalten wählen Sie Chi-Quadrat aus, um das Pearson-Chi-Quadrat, das Likelihood-Quotienten-Chi-Quadrat, den exakten Test nach Fisher und das korrigierte Chi-Quadrat nach Yates (Kontinuitätskorrektur) zu berechnen. Für 2 × 2-Tabellen wird der exakte Test nach Fisher berechnet, wenn eine Tabelle, die nicht aus fehlenden Zeilen oder Spalten

einer größeren Tabelle entstanden ist, eine Zelle mit einer erwarteten Häufigkeit von weniger als 5 enthält. Für alle anderen 2 × 2-Tabellen wird das korrigierte Chi-Quadrat nach Yates berechnet. Für Tabellen mit einer beliebigen Anzahl von Zeilen und Spalten wählen Sie Chi-Quadrat aus, um das Pearson-Chi-Quadrat und das Likelihood-Quotienten-Chi-Quadrat zu berechnen. Wenn beide Tabellenvariablen quantitativ sind, ergibt Chi-Quadrat den Zusammenhangstest linear-mit-linear.

Korrelationen. Für Tabellen, in denen sowohl Zeilen als auch Spalten geordnete Werte enthalten, ergeben die Korrelationen den Korrelationskoeffizienten nach Spearman, also Rho (nur numerische Daten). Der Korrelationskoeffizient nach Spearman ist ein Zusammenhangsmaß zwischen den Rangordnungen. Wenn beide Tabellenvariablen (Faktoren) quantitativ sind, ergibt sich unter Korrelationen der Korrelationskoeffizient nach Pearson, *r*, der ein Maß für den linearen Zusammenhang zwischen den Variablen darstellt.

Nominal. Für nominale Daten (ohne implizierte Reihenfolge wie beispielsweise katholisch, protestantisch, jüdisch) können Sie den Phi-Koeffizienten und Cramér-V, den Kontingenzkoeffizienten, Lambda (symmetrische und asymmetrische Lambdas sowie Goodman-und-Kruskal-Tau) und den Unsicherheitskoeffizienten auswählen.

- **Kontingenzkoeffizient.** Ein auf der Chi-Quadrat-Statistik basierendes Zusammenhangsmaß. Dieser Koeffizient liegt immer zwischen 0 und 1, wobei Null angibt, daß kein Zusammenhang zwischen Zeilen- und Spaltenvariable besteht und Werte nahe 1 auf einen starken Zusammenhang zwischen den Variablen hindeuten. Der maximale Wert hängt von der Anzahl der Zeilen und Spalten in der Tabelle ab.
- **Phi und Cramer-V.** Phi ist ein Zusammenhangsmaß auf der Basis der Chi-Quadrat-Statistik. Es ergibt sich als Wurzel aus dem Quotienten aus Chi-Quadrat und dem Stichprobenumfang. Cramer-V ist ebenfalls ein Zusammenhangsmaß auf der Basis der Chi-Quadrat-Statistik.
- **Lambda.** Ein Zusammenhangsmaß, welches die proportionale Fehlerreduktion wiedergibt, wenn Werte der unabhängigen Variablen zur Vorhersage der abhängigen Variablen verwendet werden. Der Wert 1 bedeutet, daß die abhängige Variable durch die unabhängige Variable vollständig vorhergesagt werden kann.

Der Wert 0 bedeutet, daß die Vorhersage der abhängigen Variablen durch die unabhängige Variable nicht unterstützt wird.

- **Unsicherheitskoeffizient.** Ein Zusammenhangsmaß der proportionalen Fehlerreduktion, wenn Werte einer Variablen zur Vorhersage von Werten der anderen Variablen verwendet werden. Ein Wert von 0,83 gibt z. B. an, daß die Kenntnis einer Variablen den Fehler bei der Vorhersage der Werte der anderen Variablen um 83% reduziert. SPSS berechnet beide Versionen des Unsicherheitskoeffizienten, die symmetrische und die asymmetrische.

Ordinal. Für Tabellen, in welchen die Zeilen und Spalten geordnete Werte enthalten, wählen Sie Gamma (nullte Ordnung für Zweifach-Tabellen und bedingt für Dreifach- bis Zehnfach-Tabellen), Kendall-Tau-b und Kendall-Tau-c aus. Zur Vorhersage von Spaltenkategorien auf der Grundlage von Zeilenkategorien wählen Sie Somers-d aus.

- **Gamma.** Ein symmetrisches Zusammenhangsmaß für zwei ordinalskalierte Variablen, dessen Wertebereich zwischen -1 und +1 liegt. Werte nahe bei -1 oder +1 weisen auf einen starken Zusammenhang zwischen den Variablen hin. Werte nahe 0 stehen für einen schwachen oder fehlenden Zusammenhang. Für Tabellen mit zwei Variablen werden Gamma-Werte nullter Ordnung, für Tabellen mit drei oder mehr Variablen bedingte Gamma-Werte angezeigt.
- **Somers-d.** Ein Zusammenhangsmaß für zwei ordinale Variablen, dessen Wertebereich zwischen -1 und +1 liegt. Werte, die betragsmäßig nahe bei 1 liegen, geben einen starken Zusammenhang zwischen den beiden Variablen an, Werte nahe 0 einen schwachen oder fehlenden Zusammenhang. Somers-d ist eine asymmetrische Erweiterung von Gamma. Der Unterschied liegt in der Einbeziehung der Anzahl von Paaren, die keine Bindungen in der unabhängigen Variablen aufweisen. SPSS berechnet auch eine symmetrische Version dieser Statistik.
- **Kendall-Tau-b.** Ein nichtparametrisches Korrelationsmaß für ordinale Variablen oder Ränge, das Bindungen berücksichtigt. Das Vorzeichen des Koeffizienten gibt die Richtung des Zusammenhangs an und sein Betrag die Stärke; dabei entsprechen betragsmäßig größere Werte einem stärkeren Zusammenhang. Die möglichen Werte liegen zwischen -1 und 1, ein Wert von -1 oder +1 ergibt sich jedoch nur aus quadratischen Tabellen.
- **Kendall-Tau-c.** Ein nichtparametrisches Zusammenhangsmaß für ordinale Variablen, das Bindungen ignoriert. Das Vorzeichen des Koeffizienten gibt die Richtung des Zusammenhangs an und sein Betrag die Stärke; dabei entsprechen betragsmäßig größere Werte einem stärkeren Zusammenhang. Die möglichen

Werte liegen zwischen -1 und +1, ein Wert von -1 oder +1 ergibt sich jedoch nur aus quadratischen Tabellen.

Nominal bezüglich Intervall. Wenn eine Variable kategorial und eine andere quantitativ ist, wählen Sie Eta aus. Die kategoriale Variable muß numerisch kodiert sein.

- **Eta.** Ein Zusammenhangsmaß, das Werte zwischen 0 und 1 annehmen kann. 0 gibt hierbei keinen Zusammenhang zwischen den Spalten- und Zeilenvariablen an, Werte nahe 1 einen hohen Grad an Zusammenhang. Eta ist geeignet, um den Zusammenhang zwischen einer intervallskalierten abhängigen Variablen (z. B.. Einkommen) und einer kategorialen unabhängigen Variablen mit einer begrenzten Anzahl von Kategorien (z. B. Geschlecht) zu beschreiben. Es werden zwei Eta-Werte berechnet: der eine behandelt die Zeilenvariable als die intervallskalierte Variable; der andere die Spaltenvariable als intervallskalierte Variable.

Kappa. Der Cohen-Kappa-Koeffizient mißt die Übereinstimmung zwischen den Beurteilungen zweier Prüfer, wenn beide dasselbe Objekt bewerten. Ein Wert von 1 bedeutet perfekte Übereinstimmung. Ein Wert von 0 bedeutet, daß die Übereinstimmung nicht über das zufallsbedingte Maß hinausgeht. Kappa ist nur für Tabellen verfügbar, in denen beide Variablen die gleiche Anzahl von Kategorien und gleiche Ausprägungen aufweisen.

Risiko. Ein Maß, das für 2 x 2-Tabellen die Stärke des Zusammenhangs zwischen dem Vorhandensein eines Faktors und dem Auftreten eines Ereignisses misst. Wenn das Konfidenzintervall für die Statistik den Wert 1 enthält, kann man nicht annehmen, dass zwischen Faktor und Ereignis ein Zusammenhang besteht. Das Quotenverhältnis (Odds Ratio) kann als Schätzer für das relative Risiko verwendet werden, wenn der Faktor selten auftritt.

McNemar. Ein nichtparametrischer Test für zwei verbundene dichotome Variablen. Prüft mit der Chi-Quadrat-Verteilung auf Änderungen bei der Reaktion. Dieser Test ist nützlich beim Feststellen von Änderungen bei der Reaktion, die auf einen experimentellen Eingriff in Vorher/Nachher-Designs zurückzuführen sind. Für größere quadratische Tabellen wird der McNemar-Bowker Test auf Symmetrie ausgegeben.

Cochran- und Mantel-Haenszel-Statistik. Cochran- und Mantel-Haenszel-Statistik können verwendet werden, um auf Unabhängigkeit zwischen einer dichotomen Faktorvariablen und einer dichotomen Response-Variable zu testen, und

zwar in Abhängigkeit von einem Kovariatenmuster, das durch eine oder mehrere Schichtvariablen (Kontrollvariablen) definiert wird. Beachten Sie, dass andere Statistiken schichtenweise berechnet werden, die Cochran- und Mantel-Haenszel-Statistik dagegen einmal für alle Schichten berechnet werden.

Kreuztabellen: Zellenanzeige

Abbildung 17-4
Dialogfeld "Kreuztabellen: Zellenanzeige"

Um Sie beim Erkennen von Mustern in den Daten zu unterstützen, die zu einem signifikanten Chi-Quadrat-Test beitragen, zeigt die Prozedur "Kreuztabellen" die erwarteten Häufigkeiten und drei Typen von Residuen (Abweichungen) an, welche die Differenz zwischen beobachteten und erwarteten Häufigkeiten messen. Jede Zelle der Tabelle kann jede Kombination von ausgewählten Häufigkeiten, Prozentzahlen und Residuen enthalten.

Häufigkeiten. Die Anzahl der Fälle, die tatsächlich beobachtet, und die Anzahl der Fälle, die erwartet werden, wenn die Zeilen- und Spaltenvariablen voneinander unabhängig sind.

Prozentwerte. Die Prozentwerte können horizontal in den Zeilen oder vertikal in den Spalten addiert werden. Der prozentuale Anteil der Gesamtanzahl der Fälle, die in einer Tabelle dargestellt werden (eine Schicht), ist ebenfalls verfügbar.

Residuen. Einfache nicht standardisierte Residuen geben die Differenz zwischen den beobachteten und erwarteten Werten wieder. Standardisierte und korrigierte standardisierte Residuen sind ebenfalls verfügbar.

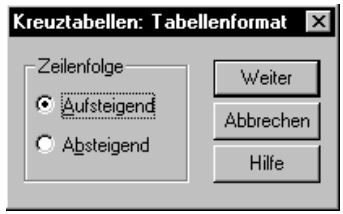
- **Nicht standardisiert.** Die Differenz zwischen einem beobachteten Wert und dem erwarteten Wert. Der erwartete Wert ist die Anzahl von Fällen, die man in einer Zelle erwarten würden, wenn kein Zusammenhang zwischen beiden Variablen bestünde. Ein positives Residuum zeigt an, dass es in der Zelle mehr Fälle gibt, als dies der Fall wäre, wenn die Zeilen- und Spaltenvariable unabhängig wären.
- **Standardisiert.** Der Quotient aus dem Residuum und einem Schätzer seines Standardfehlers. Standardisierte Residuen, auch bekannt als Pearson-Residuen, haben einen Mittelwert von 0 und eine Standardabweichung von 1.
- **Korrigiert standardisiert.** Der Quotient aus dem Residuum einer Zelle (beobachteter Wert minus erwarteter Wert) und dessen geschätztem Standardfehler. Das resultierende standardisierte Residuum wird in Einheiten der Standardabweichung über oder unter dem Mittelwert angegeben.

Nicht-ganzzahlige Gewichtungen. Bei den Zellhäufigkeiten handelt es sich normalerweise um ganzzahlige Werte, da sie für die Anzahl der Fälle in den einzelnen Zellen stehen. Wenn jedoch die Datendatei derzeit mit einer Gewichtungsvariablen mit Bruchzahlenwerten (z. B. 1,25) gewichtet ist, können die Zellhäufigkeiten ebenfalls Bruchwerte sein. Sie können die Werte vor oder nach der Berechnung der Zellhäufigkeiten abschneiden oder runden oder sowohl für die Tabellenanzeige als auch für statistische Berechnungen gebrochene Zellhäufigkeiten verwenden.

- **Anzahl in den Zellen runden.** Fallgewichte werden verwendet wie gegeben, aber die addierten Gewichte für die Zellen werden gerundet bevor Statistiken berechnet werden.
- **Anzahl in den Zellen stutzen.** Fallgewichte werden verwendet wie gegeben, aber die addierten Gewichte für die Zellen werden auf den ganzzahligen Anteil gestutzt bevor Statistiken berechnet werden.
- **Fallgewichte runden.** Fallgewichte werden gerundet bevor sie verwendet werden.
- **Fallgewichte stutzen.** Fallgewichte werden auf den ganzzahligen Anteil gestutzt bevor sie verwendet werden.
- **Keine Korrekturen.** Fallgewichte werden verwendet wie gegeben und auch nicht-ganzzahlige Zellhäufigkeiten werden verwendet. Wenn jedoch Exakte Statistiken (verfügbar mit dem Modul Exakte Tests) angefordert werden, dann werden die addierten Gewichte in den Zellen entweder auf den ganzzahligen Anteil gestutzt oder gerundet bevor die Statistiken für exakte Tests berechnet werden.

Kreuztabellen: Tabellenformat

Abbildung 17-5
Dialogfeld "Kreuztabellen: Tabellenformat"



Sie können Zeilen in aufsteigender oder absteigender Reihenfolge der Werte der Zeilenvariablen anordnen.

Kapitel
18

Zusammenfassen

Mit der Prozedur "Zusammenfassen" werden Untergruppenstatistiken für Variablen innerhalb der Kategorien einer oder mehrerer Gruppenvariablen berechnet. Alle Ebenen der Gruppenvariablen werden in die Kreuztabelle aufgenommen. Sie können wählen, in welcher Reihenfolge die Statistiken angezeigt werden. Außerdem werden Auswertungsstatistiken für jede Variable über alle Kategorien angezeigt. Die Datenwerte jeder Kategorie können aufgelistet oder unterdrückt werden. Bei umfangreichen Datensätzen haben Sie die Möglichkeit, nur die ersten *n* Fälle aufzulisten.

Beispiel. Wie hoch liegen die durchschnittlichen Verkaufszahlen eines Produkts, gegliedert nach Region und Abnehmer? Möglicherweise stellen Sie fest, daß im Westen im Durchschnitt geringfügig mehr verkauft wird als in anderen Regionen, wobei gewerbliche Kunden in der westlichen Region die wichtigsten Abnehmer sind.

Statistiken. Summe, Anzahl der Fälle, Mittelwert, Median, gruppierter Median, Standardfehler des Mittelwerts, Minimum, Maximum, Spannweite, Variablenwert der ersten Kategorie der Gruppenvariablen, Variablenwert der letzten Kategorie der Gruppenvariablen, Standardabweichung, Varianz, Kurtosis, Standardfehler der Kurtosis, Schiefe, Standardfehler der Schiefe, Prozent der Gesamtsumme, Prozent der Gesamtanzahl, Prozent der Summe in, Prozent der Anzahl in, geometrisches Mittel und harmonisches Mittel.

Daten. Die Gruppenvariablen stellen kategoriale Variablen dar, deren Werte numerisch oder kurze Strings sein können. Die Anzahl der Kategorien sollte angemessen klein gehalten werden. Den anderen Variablen müssen Ränge zugeordnet werden können.

Annahmen. Einige der möglichen Untergruppenstatistiken, wie beispielsweise Mittelwert und Standardabweichung, basieren auf der Annahme, daß eine Normalverteilung vorliegt, und sind für Variablen mit symmetrischen Verteilungen geeignet. Robuste Statistiken, wie beispielsweise Median und Spannweite,

sind für quantitative Variablen geeignet, die möglicherweise die Annahme einer Normalverteilung erfüllen.

Abbildung 18-1
Ausgabe von Zusammenfassen

Zusammenfassungen von Fällen
Durchschnittlicher Produktverkauf nach Region und Branche

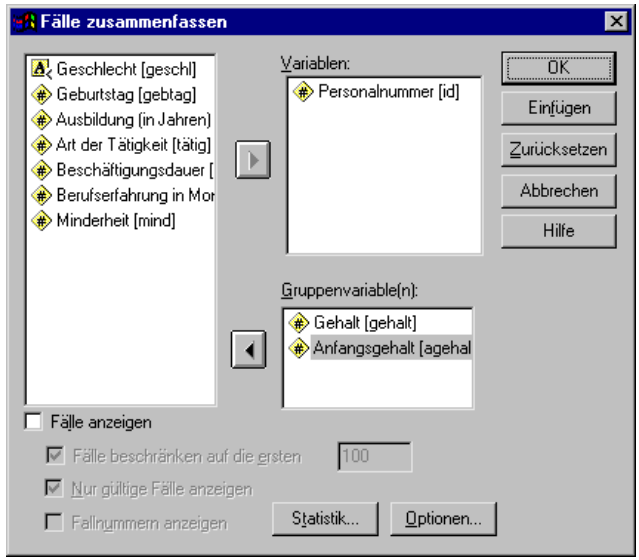
	Branche				
Region	Staat	Unter-nehmen	Bildungs-reinrichtungen	Gesund-heitswesen	Gesamt
Osten	$3.594,65	$3.953,76	$3.764,91	$3.722,32	$3.735,45
Mitte	$3.370,12	$3.268,47	$3.317,81	$3.165,11	$3.305,03
Westen	$3.552,50	$4.649,00	$4.276,25	$4.027,00	$4.079,46
Gesamt	$3.503,75	$3.727,50	$3.579,76	$3.456,93	$3.576,52

So erstellen Sie Zusammenfassungen von Fällen

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Berichte
    Fälle zusammenfassen...

Abbildung 18-2
Dialogfeld "Fälle zusammenfassen"



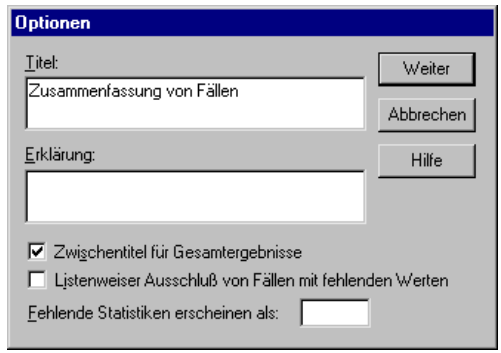
▶ Wählen Sie mindestens eine Variable aus.

Die folgenden Optionen sind verfügbar:

- Sie können eine oder mehrere Gruppenvariablen auswählen, um die Daten in Untergruppen aufzuteilen.
- Klicken Sie auf Optionen, wenn Sie den Ausgabetitel ändern, eine Erklärung unter der Ausgabe hinzufügen oder Fälle mit fehlenden Werten ausschließen möchten.
- Sie können optionale Statistiken anzeigen lassen, indem Sie auf Statistik klicken.
- Wählen Sie Fälle anzeigen, um die Fälle in jeder Untergruppe auflisten zu lassen. In der Standardeinstellung werden nur die ersten 100 Fälle in der Datei aufgelistet. Sie können den Wert für Fälle beschränken auf die ersten*n* erhöhen oder vermindern bzw. diese Option deaktivieren, um alle Fälle auflisten zu lassen.

Zusammenfassen: Optionen

Abbildung 18-3
Dialogfeld "Optionen" für "Fälle zusammenfassen"

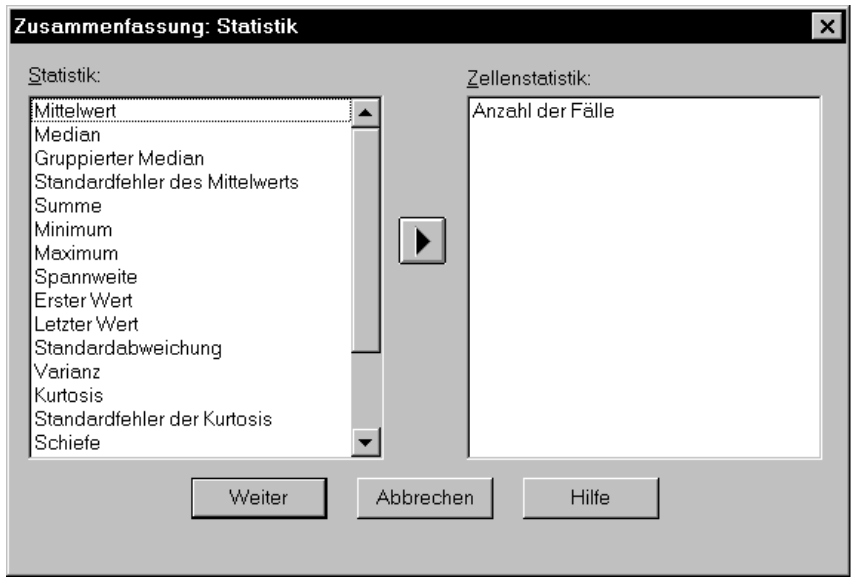


Sie können den Titel der Ausgabe ändern oder eine Erklärung hinzufügen, die unter der Ausgabetabelle angezeigt wird. Sie können den Zeilenumbruch in Titeln und Erklärungen steuern, indem Sie an die Stellen, an denen ein Zeilenumbruch durchgeführt werden soll, die Zeichen \n eingeben.

Außerdem können Sie Untertitel für Gesamtergebnisse ein- oder ausblenden sowie Fälle mit fehlenden Werten für beliebige, in der Analyse verwendete Variablen ein- oder ausschließen. Oft ist es angebracht, fehlende Fälle in der Ausgabe mit einem Punkt oder einem Sternchen zu kennzeichnen. Geben Sie ein Zeichen, eine Wortgruppe oder einen Code ein, der bei einem fehlenden Wert angezeigt werden soll, andernfalls werden fehlende Werte in der Ausgabe nicht besonders verarbeitet.

Zusammenfassung: Statistik

Abbildung 18-4
Dialogfeld "Zusammenfassung: Statistik"



Sie können mindestens eine der folgenden Untergruppen-Statistiken für die Variablen innerhalb jeder Kategorie jeder Gruppenvariablen auswählen: Summe, Anzahl der Fälle, Mittelwert, Median, gruppierter Median, Standardfehler des Mittelwerts, Minimum, Maximum, Spannweite, Variablenwert der ersten Kategorie der Gruppenvariablen, Variablenwert der letzten Kategorie der Gruppenvariablen, Standardabweichung, Varianz, Kurtosis, Standardfehler der Kurtosis, Schiefe, Standardfehler der Schiefe, Prozent der Gesamtsumme, Prozent der Gesamtanzahl, Prozent der Summe in, Prozent der Anzahl in, geometrisches Mittel und harmonisches Mittel. Die Statistiken werden in der Liste "Zellenstatistik" in derselben Reihenfolge angezeigt, in welcher sie in der Ausgabe angezeigt werden. Außerdem werden die Auswertungsstatistiken für jede Variable über alle Kategorien angezeigt.

Erster. Zeigt den ersten Datenwert in der Datendatei an.

Geometrisches Mittel. Die n-te Wurzel aus dem Produkt der Datenwerte, wobei n der Anzahl der Fälle entspricht.

Gruppierter Median. Der Median für Daten, bei denen ein Wert für ein ganzes Intervall steht. Wenn z. B. für Altersdaten alle 30er Werte als 35 erfasst werden, alle 40er Werte als 45, usw., dann ist der gruppierte Median der Median, der sich aus den erfassten Werten berechnet.

Harmonisches Mittel. Wird verwendet, um die "mittlere" Gruppengröße zu bestimmen, wenn der Stichprobenumfang in den einzelnen Gruppen unterschiedlich ist. Der harmonische Mittelwert ist gleich der Gesamtzahl der Stichproben geteilt durch die Summe der Kehrwerte der Stichprobengrößen.

Kurtosis. Ein Maß dafür, wie sich die Beobachtungen um einen zentralen Punkt gruppieren. Bei einer Normalverteilung ist der Wert der Kurtosis gleich 0. Bei positiver Kurtosis gruppieren sich die Beobachtungen dichter als bei der Normalverteilung und haben längere Flügel. Bei negativer Kurtosis gruppieren sich die Beobachtungen weniger dicht zusammen und haben kürzere Flügel.

Letzter. Hiermit wird der letzte Datenwert in der Datendatei angezeigt.

Maximum. Der größte Wert einer numerischen Variablen.

Mittelwert. Ein Lagemaß. Das arithmetische Mittel, d. h. die Summe geteilt durch die Anzahl der Fälle.

Median. Der Wert, über und unter dem jeweils die Hälfte aller Fälle liegt; das 50. Perzentil. Bei einer geraden Anzahl von Fällen ist der Median der Mittelwert der zwei mittleren Fälle, wenn diese auf- oder absteigend sortiert sind. Der Median ist ein gegenüber Ausreißern unempfindliches Lagemaß - anders als der Mittelwert, der durch wenige extrem hohe oder niedrige Werte beeinflußt wird.

Minimum. Der kleinste Wert einer numerischen Variablen.

N. Die Anzahl der Fälle (Beobachtungen oder Datensätze).

Prozent der Gesamtanzahl. Prozentsatz der Gesamtanzahl von Fällen in jeder Kategorie.

Prozent der Gesamtsumme. Prozentsatz der Gesamtsumme in jeder Kategorie.

Spannweite. Die Differenz zwischen den größten und kleinsten Werten einer numerischen Variablen; Maximalwert minus Minimalwert.

Schiefe. Ein Maß für die Asymmetrie einer Verteilung. Die Normalverteilung ist symmetrisch, ihre Schiefe hat den Wert Null. Eine Verteilung mit einer deutlichen positiven Schiefe läuft nach rechts lang aus (langer rechter Schwanz), eine Verteilung

mit einer deutlichen negativen Schiefe läuft nach links lang aus (langer linker Schwanz). Als Faustregel kann man verwenden, dass ein Schiefe-Wert, der mehr als doppelt so groß ist wie sein Standardfehler für eine Abweichung von der Symmetrie spricht.

Standardfehler der Kurtosis. Das Verhältnis der Kurtosis zu ihrem Standardfehler kann als Test für die Normalität verwendet werden (d. h. die Normalität kann zurückgewiesen werden, wenn das Verhältnis kleiner als -2 oder größer als +2 ist). Ein großer positiver Wert für die Kurtosis deutet darauf hin, daß die Schwänze der Verteilung länger sind als bei einer Normalverteilung; ein negativer Wert bedeutet, daß sie kürzer sind (etwa wie bei einer kastenförmigen, gleichförmigen Verteilung).

Standardfehler der Schiefe. Das Verhältnis der Schiefe zu ihrem Standardfehler kann für einen Test auf Normalverteilung verwendet werden (d. h. die Annahme, dass Normalverteilung vorliegt kann abgelehnt werden, wenn das Verhältnis kleiner als -2 oder größer als +2 ist). Ein großer positiver Wert für die Schiefe bedeutet, dass die Verteilung eine lange rechte Seite hat; ein extremer negativer Wert bedeutet, dass sie eine lange linke Seite hat.

Summe. Die Summe der Werte über alle Fälle mit nichtfehlenden Werten.

Varianz. Ein Maß der Streuung um den Mittelwert. Es ist gleich dem Quotienten aus der Summe der quadrierten Abweichung vom Mittelwert und der um 1 verringerten Fallanzahl. Die Varianz wird in Einheiten gemessen, die das Quadrat der Einheiten der Variablen selbst sind.

Kapitel
19

Mittelwerte

Mit der Prozedur "Mittelwerte" werden die Mittelwerte von Untergruppen und verwandte univariate Statistiken für abhängige Variablen innerhalb von Kategorien von mindestens einer unabhängigen Variablen berechnet. Wahlweise können Sie eine einfaktorielle Varianzanalyse, Eta und einen Test auf Linearität berechnen lassen.

Beispiel. Sie messen die mittlere Menge von Fett, die von drei verschiedenen Sorten Speiseöl absorbiert wird. Anschließend führen Sie eine einfaktorielle Varianzanalyse aus, um festzustellen, ob sich die Mittelwerte unterscheiden.

Statistiken. Summe, Anzahl der Fälle, Mittelwert, Median, gruppierter Median, Standardfehler des Mittelwerts, Minimum, Maximum, Spannweite, Variablenwert der ersten Kategorie der Gruppenvariablen, Variablenwert der letzten Kategorie der Gruppenvariablen, Standardabweichung, Varianz, Kurtosis, Standardfehler der Kurtosis, Schiefe, Standardfehler der Schiefe, Prozent der Gesamtsumme, Prozent der Gesamtanzahl, Prozent der Summe in, Prozent der Anzahl in, geometrisches Mittel und harmonisches Mittel. Unter Optionen stehen außerdem Varianzanalyse, Eta, Eta-Quadrat, R und R^2 zur Verfügung.

Daten. Die abhängigen Variablen sind quantitativ, die unabhängigen Variablen kategorial. Die Werte der kategorialen Variablen können numerisch oder kurze Strings sein.

Annahmen. Einige der möglichen Untergruppenstatistiken, wie beispielsweise Mittelwert und Standardabweichung, basieren auf der Annahme, daß eine Normalverteilung vorliegt, und sind für Variablen mit symmetrischen Verteilungen geeignet. Robuste Statistiken, z. B. Median, sind für quantitative Variablen geeignet, die möglicherweise die Annahme einer Normalverteilung erfüllen. Die Varianzanalyse ist gegenüber Abweichungen von der Normalverteilung robust. Allerdings sollten die Daten in jeder Zelle symmetrisch sein. Bei der Varianzanalyse wird außerdem angenommen, daß die Gruppen aus Grundgesamtheiten mit gleichen Varianzen stammen. Zum Testen dieser Annahme können Sie den Levene-Test auf

Homogenität der Varianzen verwenden. Dieser Test ist in der Prozedur "Einfaktorielle ANOVA" verfügbar.

Abbildung 19-1
Ausgabe der Mittelwerte

Bericht

FETT

ÖL			
	Erdnuß-Öl	Mittelwert	72.00
		N	6
		Standardabweichung	13.34
	Schweineschmalz	Mittelwert	85.00
		N	6
		Standardabweichung	7.77
	Mais-Öl	Mittelwert	62.00
		N	6
		Standardabweichung	8.22
	Insgesamt	Mittelwert	73.00
		N	18
		Standardabweichung	13.56

ANOVA-Tabelle

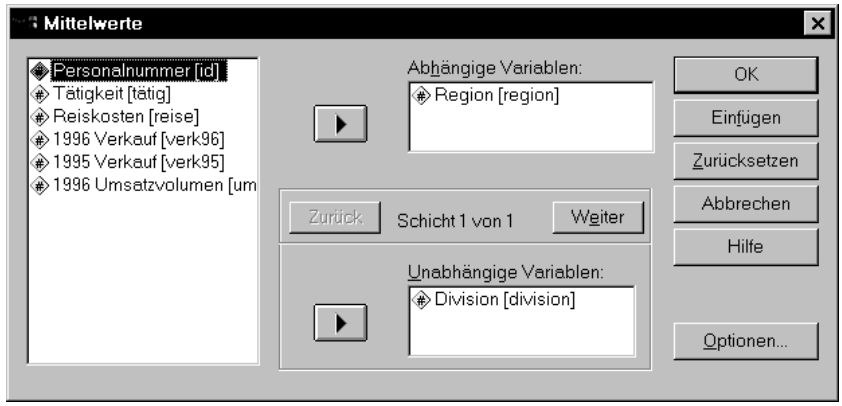
			Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
FETT * ÖL	Zwischen den Gruppen	(Kombiniert)	1596.000	2	798.000	7.824	.005
	Innerhalb der Gruppen		1530.000	15	102.000		
	Insgesamt		3126.000	17			

So berechnen Sie die Mittelwerte der Untergruppen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Mittelwerte vergleichen
    Mittelwerte...

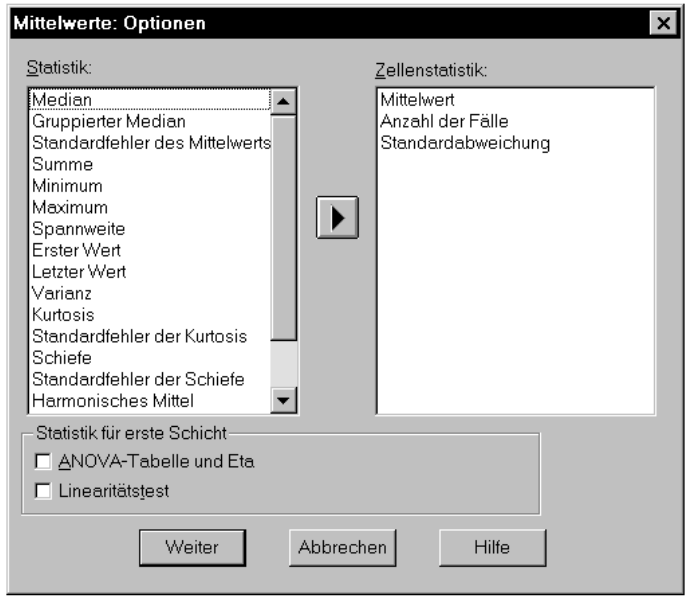
Abbildung 19-2
Dialogfeld "Mittelwerte"



- Wählen Sie eine oder mehrere abhängige Variablen aus.
- Verwenden Sie eine der folgenden Methoden, um die kategorialen unabhängigen Variablen auszuwählen:
  - Wählen Sie mindestens eine unabhängige Variable aus. Für jede unabhängige Variable werden getrennte Ergebnisse angezeigt.
  - Wählen Sie mindestens eine Schicht von unabhängigen Variablen aus. Die Stichprobe wird durch jede Schicht weiter unterteilt. Wenn es eine unabhängige Variable in Schicht 1 und eine unabhängige Variable in Schicht 2 gibt, werden die Ergebnisse nicht in einzelnen Tabellen für die unabhängigen Variablen, sondern in einer Kreuztabelle angezeigt.
- Sie können optionale Statistiken, eine Tabelle für die Varianzanalyse, Eta, Eta-Quadrat, R und R^2 berechnen lassen, indem Sie auf Optionen klicken.

Mittelwerte: Optionen

Abbildung 19-3
Dialogfeld "Mittelwerte: Optionen"



Sie können mindestens eine der folgenden Untergruppen-Statistiken für die Variablen innerhalb jeder Kategorie jeder Gruppenvariablen auswählen: Summe, Anzahl der Fälle, Mittelwert, Median, gruppierter Median, Standardfehler des Mittelwerts, Minimum, Maximum, Spannweite, Variablenwert der ersten Kategorie der Gruppenvariablen, Variablenwert der letzten Kategorie der Gruppenvariablen, Standardabweichung, Varianz, Kurtosis, Standardfehler der Kurtosis, Schiefe, Standardfehler der Schiefe, Prozent der Gesamtsumme, Prozent der Gesamtanzahl, Prozent der Summe in, Prozent der Anzahl in, geometrisches Mittel und harmonisches Mittel. Sie können die Reihenfolge ändern, in der die Statistiken für die Untergruppen berechnet werden. Die Statistiken werden in der Liste "Zellenstatistik" in derselben Reihenfolge angezeigt, in der sie in der Ausgabe angezeigt werden. Außerdem werden die Auswertungsstatistiken für jede Variable über alle Kategorien angezeigt.

Erster. Zeigt den ersten Datenwert in der Datendatei an.

Geometrisches Mittel. Die n-te Wurzel aus dem Produkt der Datenwerte, wobei n der Anzahl der Fälle entspricht.

Gruppierter Median. Der Median für Daten, bei denen ein Wert für ein ganzes Intervall steht. Wenn z. B. für Altersdaten alle 30er Werte als 35 erfasst werden, alle 40er Werte als 45, usw., dann ist der gruppierte Median der Median, der sich aus den erfassten Werten berechnet.

Harmonisches Mittel. Wird verwendet, um die "mittlere" Gruppengröße zu bestimmen, wenn der Stichprobenumfang in den einzelnen Gruppen unterschiedlich ist. Der harmonische Mittelwert ist gleich der Gesamtzahl der Stichproben geteilt durch die Summe der Kehrwerte der Stichprobengrößen.

Kurtosis. Ein Maß dafür, wie sich die Beobachtungen um einen zentralen Punkt gruppieren. Bei einer Normalverteilung ist der Wert der Kurtosis gleich 0. Bei positiver Kurtosis gruppieren sich die Beobachtungen dichter als bei der Normalverteilung und haben längere Flügel. Bei negativer Kurtosis gruppieren sich die Beobachtungen weniger dicht zusammen und haben kürzere Flügel.

Letzter. Hiermit wird der letzte Datenwert in der Datendatei angezeigt.

Maximum. Der größte Wert einer numerischen Variablen.

Mittelwert. Ein Lagemaß. Das arithmetische Mittel, d. h. die Summe geteilt durch die Anzahl der Fälle.

Median. Der Wert, über und unter dem jeweils die Hälfte aller Fälle liegt; das 50. Perzentil. Bei einer geraden Anzahl von Fällen ist der Median der Mittelwert der zwei mittleren Fälle, wenn diese auf- oder absteigend sortiert sind. Der Median ist ein gegenüber Ausreißern unempfindliches Lagemaß - anders als der Mittelwert, der durch wenige extrem hohe oder niedrige Werte beeinflußt wird.

Minimum. Der kleinste Wert einer numerischen Variablen.

N. Die Anzahl der Fälle (Beobachtungen oder Datensätze).

Prozent der Gesamtanzahl. Prozentsatz der Gesamtanzahl von Fällen in jeder Kategorie.

Prozent der Gesamtsumme. Prozentsatz der Gesamtsumme in jeder Kategorie.

Spannweite. Die Differenz zwischen den größten und kleinsten Werten einer numerischen Variablen; Maximalwert minus Minimalwert.

Schiefe. Ein Maß für die Asymmetrie einer Verteilung. Die Normalverteilung ist symmetrisch, ihre Schiefe hat den Wert Null. Eine Verteilung mit einer deutlichen positiven Schiefe läuft nach rechts lang aus (langer rechter Schwanz), eine Verteilung mit einer deutlichen negativen Schiefe läuft nach links lang aus (langer linker Schwanz). Als Faustregel kann man verwenden, dass ein Schiefe-Wert, der mehr als doppelt so groß ist wie sein Standardfehler für eine Abweichung von der Symmetrie spricht.

Standardfehler der Kurtosis. Das Verhältnis der Kurtosis zu ihrem Standardfehler kann als Test für die Normalität verwendet werden (d. h. die Normalität kann zurückgewiesen werden, wenn das Verhältnis kleiner als -2 oder größer als +2 ist). Ein großer positiver Wert für die Kurtosis deutet darauf hin, daß die Schwänze der Verteilung länger sind als bei einer Normalverteilung; ein negativer Wert bedeutet, daß sie kürzer sind (etwa wie bei einer kastenförmigen, gleichförmigen Verteilung).

Standardfehler der Schiefe. Das Verhältnis der Schiefe zu ihrem Standardfehler kann für einen Test auf Normalverteilung verwendet werden (d. h. die Annahme, dass Normalverteilung vorliegt kann abgelehnt werden, wenn das Verhältnis kleiner als -2 oder größer als +2 ist). Ein großer positiver Wert für die Schiefe bedeutet, dass die Verteilung eine lange rechte Seite hat; ein extremer negativer Wert bedeutet, dass sie eine lange linke Seite hat.

Summe. Die Summe der Werte über alle Fälle mit nichtfehlenden Werten.

Varianz. Ein Maß der Streuung um den Mittelwert. Es ist gleich dem Quotienten aus der Summe der quadrierten Abweichung vom Mittelwert und der um 1 verringerten Fallanzahl. Die Varianz wird in Einheiten gemessen, die das Quadrat der Einheiten der Variablen selbst sind.

Statistik für erste Schicht

ANOVA-Tabelle und Eta. Zeigt eine Tabelle für eine einfaktorielle Varianzanalyse an und berechnet Eta und Eta-Quadrat (Zusammenhangsmaße) für jede unabhängige Variable in der ersten Schicht.

Linearitätstest. Berechnet für lineare und nichtlineare Komponenten die Quadratsummen, die Freiheitsgrade und das Mittel der Quadrate, sowie den F-Wert, R und R-Quadrat. Die Berechnungen für Linearität werden nicht durchgeführt, wenn die unabhängige Variable eine kurze String-Variable ist.

Kapitel
20

OLAP-Würfel

Mit der Prozedur “OLAP-Würfel” (Online Analytical Processing) werden Gesamtwerte, Mittelwerte und andere univariate Statistiken für stetige Auswertungsvariablen innerhalb der Kategorien von mindestens einer kategorialen Gruppenvariablen berechnet. Für jede Kategorie der Gruppenvariablen wird eine separate Schicht erstellt.

Beispiel. Durchschnittlicher und gesamter Umsatz für verschiedene Regionen und Produktlinien innerhalb einer Region.

Statistiken. Summe, Anzahl der Fälle, Mittelwert, Median, Gruppierter Median, Standardfehler des Mittelwerts, Minimum, Maximum, Spannweite, Variablenwert der ersten Kategorie der Gruppenvariablen, Variablenwert der letzten Kategorie der Gruppenvariablen, Standardabweichung, Varianz, Kurtosis, Standardfehler der Kurtosis, Schiefe, Standardfehler der Schiefe, Prozentsatz der gesamten Fälle, Prozentsatz der Gesamtsumme, Prozentsatz der gesamten Fälle innerhalb der Gruppenvariablen, Prozentsatz der Gesamtsumme innerhalb der Gruppenvariablen, geometrisches Mittel und harmonisches Mittel.

Daten. Die Auswertungsvariablen sind quantitativ (stetige Variablen, die auf einer Intervall- oder Verhältnisskala gemessen werden) und die Gruppenvariablen kategorial. Die Werte der kategorialen Variablen können numerisch oder kurze Strings sein.

Annahmen. Einige der möglichen Untergruppenstatistiken, wie beispielsweise Mittelwert und Standardabweichung, basieren auf der Annahme, daß eine Normalverteilung vorliegt, und sind für Variablen mit symmetrischen Verteilungen geeignet. Robuste Statistiken, wie z. B. Median und Spannweite, sind für quantitative Variablen geeignet, die möglicherweise die Annahme einer Normalverteilung erfüllen.

Abbildung 20-1
Ausgabe von "OLAP-Würfel"

Umsätze 1996 nach Abteilung und Region

Abteilung: Gesampt
Region: Gesampt

Summe	$145.038,250
Mittelwert	$371,893
Median	$307,500
Standardabweichung	$171,311

Umsätze 1996 nach Abteilung und Region

Abteilung: Kundenprodukte
Region: Osten

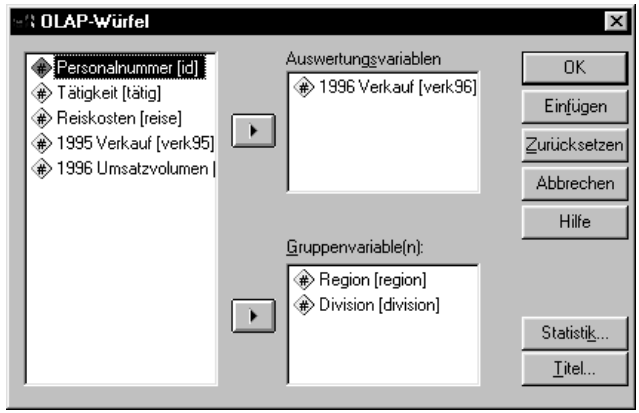
Summe	$18.548,100
Mittelwert	$289.814,06
Median	$273.600,00
Standardabweichung	$80.674,66

So erstellen Sie OLAP-Würfel

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Berichte
  OLAP-Würfel...

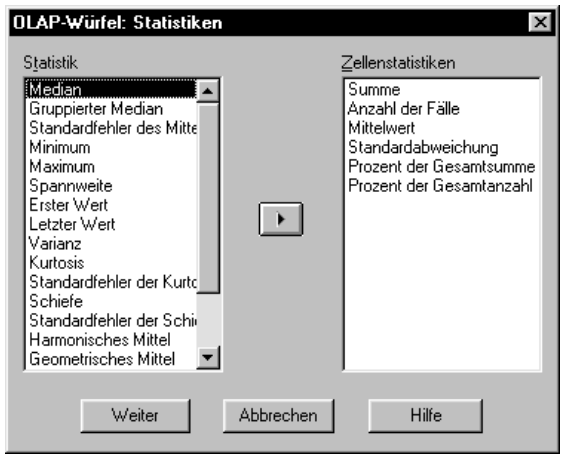
Abbildung 20-2
Dialogfeld "OLAP-Würfel"



- Wählen Sie mindestens eine stetige Auswertungsvariable aus.
- Wählen Sie mindestens eine kategoriale Gruppenvariable aus.
- Die folgenden Optionen sind verfügbar:
  - Sie können verschiedene Auswertungsstatistiken auswählen, indem Sie auf Statistiken klicken. Sie müssen mindestens eine Gruppenvariable auswählen, bevor Sie die Auswertungsstatistiken auswählen können.
  - Sie können die Differenzen zwischen Variablenpaaren und Gruppenpaaren berechnen lassen, die durch eine Gruppenvariable definiert sind, indem Sie auf Differenzen klicken.
  - Sie können Titel für benutzerdefinierte Tabellen erstellen, indem Sie auf Titel klicken.

OLAP-Würfel: Statistiken

Abbildung 20-3
Dialogfeld "OLAP-Würfel: Statistiken"



Sie können eine oder mehrere der folgenden Untergruppen-Statistiken für die Auswertungsvariablen in jeder Kategorie aller Gruppenvariablen auswählen: Summe, Anzahl der Fälle, Mittelwert, Median, Gruppierter Median, Standardfehler des Mittelwerts, Minimum, Maximum, Spannweite, Variablenwert der ersten Kategorie der Gruppenvariablen, Variablenwert der letzten Kategorie der Gruppenvariablen, Standardabweichung, Varianz, Kurtosis, Standardfehler der Kurtosis, Schiefe, Standardfehler der Schiefe, Prozentsatz der gesamten Fälle, Prozentsatz der Gesamtsumme, Prozentsatz der gesamten Fälle innerhalb der Gruppenvariablen, Prozentsatz der Gesamtsumme innerhalb der Gruppenvariablen, geometrisches Mittel und harmonisches Mittel.

Sie können die Reihenfolge ändern, in der die Statistiken für die Untergruppen berechnet werden. Die Statistiken werden in der Liste "Zellenstatistik" in derselben Reihenfolge angezeigt, in der sie in der Ausgabe angezeigt werden. Außerdem werden die Auswertungsstatistiken für jede Variable über alle Kategorien angezeigt.

Erster. Zeigt den ersten Datenwert in der Datendatei an.

Geometrisches Mittel. Die n-te Wurzel aus dem Produkt der Datenwerte, wobei n der Anzahl der Fälle entspricht.

Gruppierter Median. Der Median für Daten, bei denen ein Wert für ein ganzes Intervall steht. Wenn z. B. für Altersdaten alle 30er Werte als 35 erfasst werden, alle 40er Werte als 45, usw., dann ist der gruppierte Median der Median, der sich aus den erfassten Werten berechnet.

Harmonisches Mittel. Wird verwendet, um die "mittlere" Gruppengröße zu bestimmen, wenn der Stichprobenumfang in den einzelnen Gruppen unterschiedlich ist. Der harmonische Mittelwert ist gleich der Gesamtzahl der Stichproben geteilt durch die Summe der Kehrwerte der Stichprobengrößen.

Kurtosis. Ein Maß dafür, wie sich die Beobachtungen um einen zentralen Punkt gruppieren. Bei einer Normalverteilung ist der Wert der Kurtosis gleich 0. Bei positiver Kurtosis gruppieren sich die Beobachtungen dichter als bei der Normalverteilung und haben längere Flügel. Bei negativer Kurtosis gruppieren sich die Beobachtungen weniger dicht zusammen und haben kürzere Flügel.

Letzter. Hiermit wird der letzte Datenwert in der Datendatei angezeigt.

Maximum. Der größte Wert einer numerischen Variablen.

Mittelwert. Ein Lagemaß. Das arithmetische Mittel, d. h. die Summe geteilt durch die Anzahl der Fälle.

Median. Der Wert, über und unter dem jeweils die Hälfte aller Fälle liegt; das 50. Perzentil. Bei einer geraden Anzahl von Fällen ist der Median der Mittelwert der zwei mittleren Fälle, wenn diese auf- oder absteigend sortiert sind. Der Median ist ein gegenüber Ausreißern unempfindliches Lagemaß - anders als der Mittelwert, der durch wenige extrem hohe oder niedrige Werte beeinflußt wird.

Minimum. Der kleinste Wert einer numerischen Variablen.

N. Die Anzahl der Fälle (Beobachtungen oder Datensätze).

Prozent der Anzahl in. Prozentsatz der Gesamtanzahl von Fällen für die angegebene Gruppenvariable in den Kategorien der anderen Gruppenvariablen. Wenn nur eine Gruppenvariable vorhanden ist, ist dieser Wert gleich dem Prozentsatz der Gesamtanzahl von Fällen.

Prozent der Summe in. Prozentsatz der Summe für die angegebene Gruppenvariable in den Kategorien der anderen Gruppenvariablen. Wenn nur eine Gruppenvariable vorhanden ist, ist dieser Wert gleich dem Prozentsatz der Gesamtsumme.

Prozent der Gesamtanzahl. Prozentsatz der Gesamtanzahl von Fällen in jeder Kategorie.

Prozent der Gesamtsumme. Prozentsatz der Gesamtsumme in jeder Kategorie.

Spannweite. Die Differenz zwischen den größten und kleinsten Werten einer numerischen Variablen; Maximalwert minus Minimalwert.

Schiefe. Ein Maß für die Asymmetrie einer Verteilung. Die Normalverteilung ist symmetrisch, ihre Schiefe hat den Wert Null. Eine Verteilung mit einer deutlichen positiven Schiefe läuft nach rechts lang aus (langer rechter Schwanz), eine Verteilung mit einer deutlichen negativen Schiefe läuft nach links lang aus (langer linker Schwanz). Als Faustregel kann man verwenden, dass ein Schiefe-Wert, der mehr als doppelt so groß ist wie sein Standardfehler für eine Abweichung von der Symmetrie spricht.

Standardfehler der Kurtosis. Das Verhältnis der Kurtosis zu ihrem Standardfehler kann als Test für die Normalität verwendet werden (d. h. die Normalität kann zurückgewiesen werden, wenn das Verhältnis kleiner als -2 oder größer als +2 ist). Ein großer positiver Wert für die Kurtosis deutet darauf hin, daß die Schwänze der Verteilung länger sind als bei einer Normalverteilung; ein negativer Wert bedeutet, daß sie kürzer sind (etwa wie bei einer kastenförmigen, gleichförmigen Verteilung).

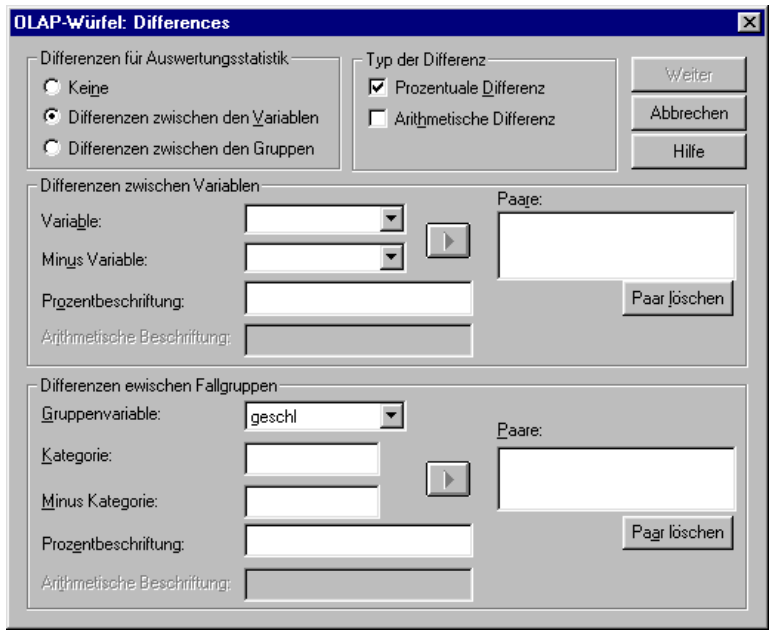
Standardfehler der Schiefe. Das Verhältnis der Schiefe zu ihrem Standardfehler kann für einen Test auf Normalverteilung verwendet werden (d. h. die Annahme, dass Normalverteilung vorliegt kann abgelehnt werden, wenn das Verhältnis kleiner als -2 oder größer als +2 ist). Ein großer positiver Wert für die Schiefe bedeutet, dass die Verteilung eine lange rechte Seite hat; ein extremer negativer Wert bedeutet, dass sie eine lange linke Seite hat.

Summe. Die Summe der Werte über alle Fälle mit nichtfehlenden Werten.

Varianz. Ein Maß der Streuung um den Mittelwert. Es ist gleich dem Quotienten aus der Summe der quadrierten Abweichung vom Mittelwert und der um 1 verringerten Fallanzahl. Die Varianz wird in Einheiten gemessen, die das Quadrat der Einheiten der Variablen selbst sind.

OLAP-Würfel: Differenzen

Abbildung 20-4
Dialogfeld "OLAP-Würfel: Differenzen"



In diesem Dialogfeld können Sie prozentuale und arithmetische Differenzen zwischen Auswertungsvariablen oder zwischen Gruppen berechnen lassen, die durch eine Gruppenvariable definiert sind. Die Differenzen werden für alle Maße berechnet, die im Dialogfeld "OLAP-Würfel: Statistiken" ausgewählt wurden.

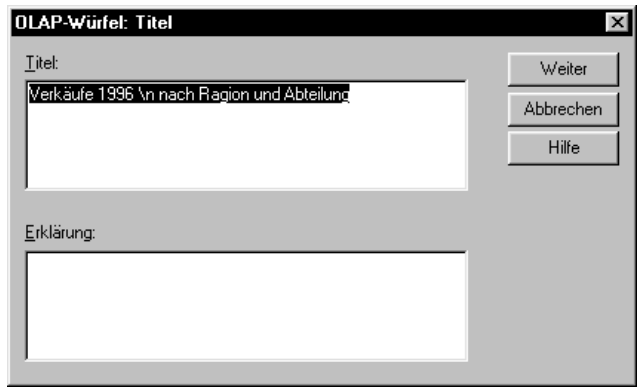
Differenzen zwischen den Variablen. Hiermit werden die Differenzen zwischen Variablenpaaren berechnet. Die Werte der Auswertungsstatistik für die zweite Variable (die Minusvariable) in jedem Paar werden von den Werten der Auswertungsstatistik für die erste Variable im Paar subtrahiert. Bei prozentualen Differenzen wird der Wert der Auswertungsvariable für die Minusvariable als Nenner verwendet. Sie müssen mindestens zwei Auswertungsvariablen im Hauptdialogfeld auswählen, bevor Sie die Differenzen zwischen den Variablen angeben können.

Differenzen zwischen Fallgruppen. Hiermit werden die Differenzen zwischen Gruppenpaaren berechnet, die durch eine Gruppenvariable definiert sind. Die Werte der Auswertungsstatistik für die zweite Kategorie (die Minuskategorie) in jedem Paar

werden von den Werten der Auswertungsstatistik für die erste Kategorie im Paar subtrahiert. Bei prozentualen Differenzen wird der Wert der Auswertungsstatistik für die Minuskategorie als Nenner verwendet. Sie müssen mindestens eine Gruppenvariable im Hauptdialogfeld auswählen, bevor Sie die Differenzen zwischen den Gruppen angeben können.

OLAP-Würfel: Titel

Abbildung 20-5
Dialogfeld "OLAP-Würfel: Titel"



Sie können den Titel der Ausgabe ändern oder eine Erklärung hinzufügen, die unter der Ausgabetabelle angezeigt wird. Sie können auch den Zeilenumbruch in Titeln und Erklärungen selbst bestimmen, indem Sie an der gewünschten Stelle im Text die Zeichenfolge \n eingeben.

Kapitel
21

T-Tests

Es sind drei Typen von *T*-Tests verfügbar:

T-Test bei unabhängigen Stichproben (T-Test bei zwei Stichproben). Vergleicht die Mittelwerte einer Variablen für zwei Fallgruppen. Für jede Gruppe sind beschreibende Statistiken und der Levene-Test auf Gleichheit der Varianzen sowie *t*-Werte für gleiche und verschiedene Varianzen und ein 95%-Konfidenzintervall für die Differenz der Mittelwerte verfügbar.

T-Test bei gepaarten Stichproben (T-Test für abhängige Variablen). Vergleicht den Mittelwert von zwei Variablen für eine einzelne Gruppe. Dieser Test ist auch für Studien mit zugeordneten Paaren oder Fallkontrolle geeignet. Die Ausgabe enthält deskriptive Statistiken für die Testvariablen, die Korrelationen zwischen den Variablen, deskriptive Statistiken für die gepaarten Differenzen, den *T*-Test und ein 95%-Konfidenzintervall.

T-Test bei einer Stichprobe. Vergleicht den Mittelwert einer Variablen mit einem bekannten oder angenommenen Wert. Neben dem *T*-Test werden deskriptive Statistiken für die Testvariablen angezeigt. In der Standardeinstellung wird unter anderem ein 95%-Konfidenzintervall für die Differenz zwischen dem Mittelwert der Testvariablen und dem angenommenen Testwert ausgegeben.

T-Test bei unabhängigen Stichproben

Im T-Test bei unabhängigen Stichproben werden die Mittelwerte von zwei Fallgruppen verglichen. Im Idealfall sollten die Subjekte bei diesem Test zufällig zwei Gruppen zugeordnet werden, so daß Unterschiede bei den Antworten lediglich auf die Behandlung (bzw. Nichtbehandlung) und keine sonstigen Faktoren zurückzuführen sind. Dies ist nicht der Fall, wenn Sie die Durchschnittseinkommen von Männern und Frauen vergleichen. Die jeweiligen Personen sind nicht zufällig auf die Gruppen “männlich” oder “weiblich” verteilt. In solchen Situationen müssen Sie sicherstellen,

daß signifikante Differenzen der Mittelwerte nicht durch Abweichungen bei anderen Faktoren verborgen oder verstärkt werden. Unterschiede im Durchschnittseinkommen können auch durch Faktoren wie den Bildungsstand beeinflußt werden (nicht nur durch das Geschlecht).

Beispiel. Patienten mit hohem Blutdruck werden zufällig auf eine Kontrollgruppe und eine Versuchsgruppe verteilt. Die Patienten in der Kontrollgruppe erhalten ein Plazebo. Die Patienten der Versuchsgruppe erhalten ein neues Medikament, dessen blutdrucksenkende Wirkung erprobt werden soll. Nach zweimonatiger Behandlung wird der *T*-Test bei zwei Stichproben angewandt, um den durchschnittlichen Blutdruck der Personen in der Kontrollgruppe mit dem der Personen aus der Versuchsgruppe zu vergleichen. Bei jedem Patienten wird eine Messung vorgenommen, und er gehört zu jeweils einer (1) Gruppe.

Statistiken. Für jede Variable: Stichprobengröße, Mittelwert, Standardabweichung und Standardfehler des Mittelwerts. Für die Differenz der Mittelwerte: Mittelwert, Standardfehler und Konfidenzintervall. (Sie können das Konfidenzniveau bestimmen.) Tests: Levene-Test auf Gleichheit der Varianzen sowie *T*-Tests auf Gleichheit der Mittelwerte bei gemeinsamen und separaten Varianzen.

Daten. Die Werte der untersuchten quantitativen Variablen müssen in einer einzelnen Spalte in der Datendatei vorliegen. Zum Aufteilen der Fälle in zwei Gruppen verwendet SPSS eine Gruppenvariable mit zwei Werten. Die Gruppenvariable kann numerische Werte (wie zum Beispiel 1 und 2, oder 6,25 und 12,5) oder kurze Strings (beispielsweise *Ja* und *Nein*) enthalten. Alternativ können Sie eine quantitative Variable wie z. B. *Alter* verwenden und die Fälle durch Angabe eines Trennwerts aufteilen (der Trennwert 21 teilt *Alter* in eine Gruppe “unter 21” und eine “21 und darüber”).

Annahmen. Für den *T*-Test auf Gleichheit der Varianzen sollten die Beobachtungen unabhängige Zufallsstichproben aus Normalverteilungen mit derselben Varianz der Grundgesamtheit sein. Für den *T*-Test auf Ungleichheit der Varianzen sollten die Beobachtungen unabhängige Zufallsstichproben aus Normalverteilungen sein. Der *T*-Test mit zwei Stichproben ist relativ robust gegenüber Abweichungen von der Normalverteilung. Achten Sie bei der grafischen Überprüfung von Verteilungen darauf, daß diese symmetrisch sind und keine Ausreißer enthalten.

Abbildung 21-1
Ausgabe des T-Tests bei unabhängigen Stichproben

Group Statistics

			N	Mean	Std. Deviation	Std. Error Mean
Blood pressure	Treatment	placebo	10	142.50	17.04	5.39
		new_drug	10	116.40	13.62	4.31

Independent Samples Test

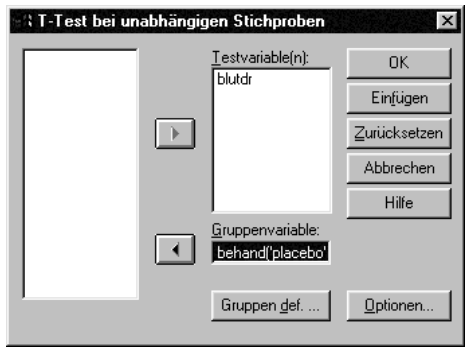
		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Mean	
		F	Significance	t	df	Significance (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Blood pressure	Equal variances assumed	.134	.719	3.783	18	.001	26.10	6.90	11.61	40.59
	Equal variances not assumed			3.783	17.163	.001	26.10	6.90	11.56	40.64

So lassen Sie einen T-Test bei unabhängigen Stichproben berechnen

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Mittelwerte vergleichen
    T-Test bei unabhängigen Stichproben...

Abbildung 21-2
Dialogfeld "T-Test bei unabhängigen Stichproben"



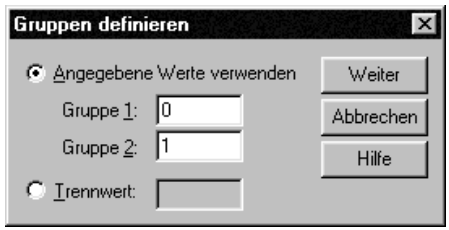
▶ Wählen Sie mindestens eine quantitative Testvariable. Für jede Variable wird ein separater *T*-Test berechnet.

- Wählen Sie eine einzelne Gruppenvariable aus, und klicken Sie dann auf Gruppen def. , um zwei Codes für die zu vergleichenden Gruppen anzugeben.

- Zusätzlich können Sie auf Optionen klicken, um die Behandlung fehlender Daten und das Niveau des Konfidenzintervalls festzulegen.

T-Test bei unabhängigen Stichproben: Gruppen definieren

Abbildung 21-3
Dialogfeld "Gruppen definieren" für numerische Variablen



Definieren Sie bei numerischen Gruppenvariablen die zwei Gruppen für den *T*-Test, indem Sie zwei Werte oder einen Trennwert angeben:

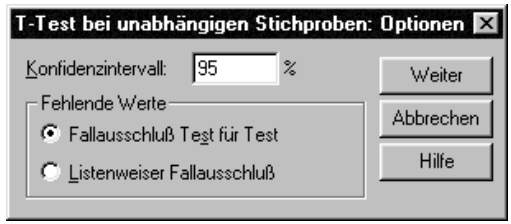
- **Angegebene Werte verwenden.** Geben Sie einen Wert für Gruppe 1 und einen weiteren Wert für Gruppe 2 ein. Fälle mit anderen Werten werden aus der Analyse ausgeschlossen. Zahlen müssen nicht ganzzahlig sein (so sind beispielsweise 6,25 und 12,5 gültige Werte).
- **Trennwert.** Geben Sie eine Zahl ein, welche die Werte der Gruppenvariablen in zwei Mengen aufteilt. Alle Fälle mit Werten, die kleiner als der Trennwert sind, bilden eine Gruppe. Die Fälle mit Werten größer oder gleich dem Trennwert bilden die andere Gruppe.

Abbildung 21-4
Dialogfeld "Gruppen definieren" für String-Variablen

Bei Gruppenvariablen mit kurzen Strings geben Sie einen String für Gruppe 1 und einen anderen für Gruppe 2 ein, beispielweise *Ja* und *Nein*. Fälle mit anderen Strings werden von der Analyse ausgeschlossen.

T-Tests bei unabhängigen Stichproben: Optionen

Abbildung 21-5
Dialogfeld "T-Test bei unabhängigen Stichproben: Optionen"



Konfidenzintervall. In der Standardeinstellung wird ein 95%-Konfidenzintervall für die Differenz der Mittelwerte angezeigt. Geben Sie einen Wert zwischen 1 und 99 ein, um ein anderes Konfidenzniveau festzulegen.

Fehlende Werte. Wenn Sie mehrere Variablen testen und bei einer oder mehreren Variablen Daten fehlen, können Sie bestimmen, welche Fälle einzuschließen (oder auszuschließen) sind.

- **Fallausschluß Test für Test.** Bei jedem *T*-Test werden alle Fälle verwendet, für die gültige Daten für die getestete Variable vorliegen. Die Stichprobengröße kann von Test zu Test unterschiedlich ausfallen.
- **Listenweiser Fallausschluß.** Jeder *T*-Test verwendet nur Fälle mit gültigen Daten für alle in den angeforderten *T*-Tests verwendeten Variablen. Die Stichprobengröße bleibt bei allen Tests konstant.

T-Test bei gepaarten Stichproben

Mit der Prozedur "T-Test bei gepaarten Stichproben" werden die Mittelwerte zweier Variablen für eine einzelne Gruppe verglichen. Diese Prozedur berechnet für jeden Fall die Differenzen zwischen den Werten der zwei Variablen und überprüft, ob der Durchschnitt von 0 abweicht.

Beispiel. In einer Studie über Bluthochdruck wird der Blutdruck aller Patienten zu Beginn der Studie und nach der Behandlung gemessen. Daher gibt es für jede Testperson zwei Meßwerte, die auch als *Vorher*- und *Nachher*-Messung bezeichnet werden. Dieser Test kann auch bei Studien mit zugeordneten Paaren bzw. mit Fallkontrolle verwendet werden. Hierbei enthält jeder Datensatz der Datendatei die Reaktion des Patienten und die von der zugehörigen Kontroll-Testperson. In einer Blutdruckstudie könnten den Patienten die Kontrollpersonen nach Alter zugeordnet werden (einem 75-jährigen Patienten ein 75-jähriges Mitglied der Kontrollgruppe).

Statistiken. Für jede Variable: Mittelwert, Stichprobengröße, Standardabweichung und Standardfehler des Mittelwerts. Für jedes Variablenpaar: Korrelation, durchschnittliche Differenz der Mittelwerte, *T*-Test und Konfidenzintervall für die Differenz der Mittelwerte. (Sie können das Konfidenzniveau festlegen.) Standardabweichung und Standardfehler der Differenz der Mittelwerte.

Daten. Legen Sie für jeden gepaarten Test zwei Variablen fest, die auf Intervall-Messniveau oder Verhältnis-Messniveau quantitativ sein müssen. In einer Studie mit zugeordneten Paaren bzw. mit Fallkontrolle müssen die Reaktionen jedes Testsubjektes und dessen zugeordneten Kontrollsubjektes im selben Fall der Datendatei enthalten sein.

Annahmen. Die Beobachtungen für jedes Paar müssen unter gleichen Bedingungen vorgenommen werden. Die Differenzen der Mittelwerte müssen normalverteilt sein. Die Varianzen jeder Variablen können gleich oder ungleich sein.

Abbildung 21-6
Ausgabe des T-Tests bei gepaarten Stichproben

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	After treatment	116.40	10	13.62	4.31
	Before treatment	142.50	10	17.04	5.39

Test bei gepaarten Stichproben

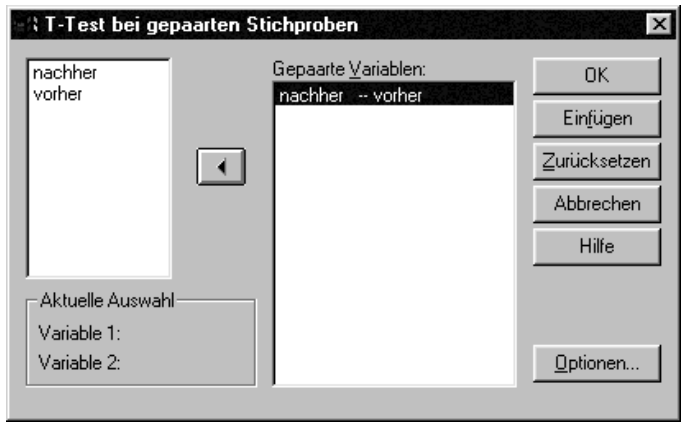
		Gepaarte Differenzen					T	df	Sig. (2-seitig)
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz Untere	Obere			
Paaren	Nach der Behandlung - Vor der Behandlung	-26.10	19.59	6.19	-40.11	-12.09	-4.214	9	.002

So lassen Sie einen T-Test bei gepaarten Stichproben berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Mittelwerte vergleichen
    T-Test bei gepaarten Stichproben...

Abbildung 21-7
Dialogfeld "T-Test bei gepaarten Stichproben"



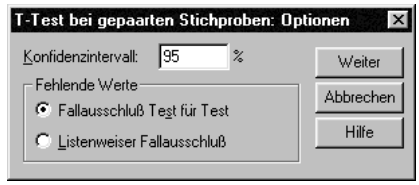
▶ Wählen Sie wie folgt ein Variablenpaar aus:

- Klicken Sie auf beide Variablen. Die erste Variable wird im Gruppenfeld "Aktuelle Auswahl" als *Variable 1*, die zweite Variable als *Variable 2* angezeigt.
- Klicken Sie auf die Pfeilschaltfläche, um das Paar in die Liste "Gepaarte Variablen" zu verschieben. Sie können mehrere Variablenpaare auswählen. Um ein Variablenpaar aus der Analyse zu entfernen, wählen Sie in der Liste "Gepaarte Variablen" ein Paar aus. Klicken Sie dann auf die Pfeilschaltfläche.

▶ Zusätzlich können Sie auf Optionen klicken, um die Behandlung fehlender Daten und das Niveau des Konfidenzintervalls festzulegen.

T-Test bei gepaarten Stichproben: Optionen

Abbildung 21-8
Dialogfeld "T-Test bei gepaarten Stichproben: Optionen"



Konfidenzintervall. In der Standardeinstellung wird ein 95%-Konfidenzintervall für die Differenz der Mittelwerte angezeigt. Geben Sie einen Wert zwischen 1 und 99 ein, um ein anderes Konfidenzniveau festzulegen.

Fehlende Werte. Wenn Sie mehrere Variablen testen und bei einer oder mehreren Variablen Daten fehlen, können Sie bestimmen, welche Fälle einzuschließen (oder auszuschließen) sind:

- **Fallausschluß Test für Test.** Bei jedem *T*-Test werden alle Fälle mit gültigen Daten für die getesteten Variablenpaare verwendet. Die Stichprobengröße kann von Test zu Test unterschiedlich ausfallen.
- **Listenweiser Fallausschluß.** Bei jedem *T*-Test werden nur Fälle mit gültigen Daten für alle getesteten Variablenpaare verwendet. Die Stichprobengröße bleibt bei allen Tests konstant.

T-Test bei einer Stichprobe

Die Prozedur "T-Test bei einer Stichprobe" prüft, ob der Mittelwert einer einzelnen Variablen von einer angegebenen Konstanten abweicht.

Beispiel. Ein Forscher könnte testen, ob der durchschnittliche IQ-Wert einer Gruppe von Studenten von 100 abweicht. Ein Hersteller von Getreideprodukten könnte stichprobenartig Packungen aus der Produktion entnehmen und prüfen, ob das Durchschnittsgewicht der Stichproben von 500 Gramm auf dem 95%-Konfidenzniveau abweicht.

Statistiken. Für jede Testvariable: Mittelwert, Standardabweichung und Standardfehler der Differenz der Mittelwerte. Außerdem die durchschnittliche Differenz zwischen jedem Datenwert und dem angenommenen Testwert, ein *T*-Test,

der prüft, ob diese Differenz null beträgt, und ein Konfidenzintervall für diese Differenz. (Sie können das Konfidenzniveau festlegen.)

Daten. Um die Werte einer quantitativen Variablen mit einem angenommenen Testwert zu vergleichen, wählen Sie eine quantitative Variable aus, und geben Sie einen angenommenen Testwert ein.

Annahmen. Bei diesem Test wird von einer Normalverteilung ausgegangen; er ist jedoch recht robust gegenüber Abweichungen von dieser Verteilung.

Abbildung 21-9
Ausgabe für den T-Test bei einer Stichprobe

One-Sample Statistics

	IQ
N	15
Mean	109.33
Std. Deviation	12.03
Std. Error Mean	3.11

Rows and columns have been transposed.

One-Sample Test

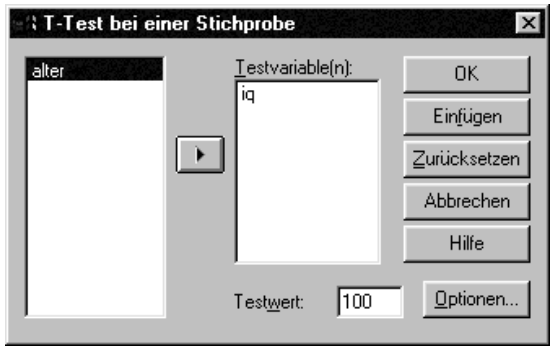
	Test Value = 100					
					95% Confidence Interval of the Difference	
	t	df	Significance (2-tailed)	Mean Difference	Lower	Upper
IQ	3.005	14	.009	9.33	2.67	15.99

So lassen Sie den T-Test bei einer Stichprobe berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Mittelwerte vergleichen
    T-Test bei einer Stichprobe…

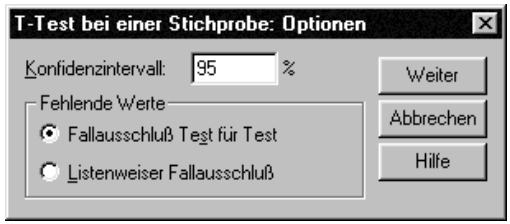
Abbildung 21-10
Dialogfeld "T-Test bei einer Stichprobe"



- Wählen Sie eine oder mehrere Variablen aus, die mit demselben angenommenen Wert verglichen werden sollen.
- Geben Sie einen numerischen Testwert ein, mit dem jeder Stichprobenmittelwert verglichen werden soll.
- Zusätzlich können Sie auf Optionen klicken, um die Behandlung fehlender Daten und das Niveau des Konfidenzintervalls festzulegen.

T-Test bei einer Stichprobe: Optionen

Abbildung 21-11
Dialogfeld "T-Test bei einer Stichprobe: Optionen"



Konfidenzintervall. In der Standardeinstellung wird ein 95%-Konfidenzintervall für die Differenz zwischen dem Mittelwert und dem angenommenen Testwert angezeigt. Geben Sie einen Wert zwischen 1 und 99 ein, um ein anderes Konfidenzniveau festzulegen.

Fehlende Werte. Wenn Sie mehrere Variablen testen und bei einer oder mehreren Variablen Daten fehlen, können Sie bestimmen, welche Fälle einzuschließen (oder auszuschließen) sind.

- **Fallausschluß Test für Test.** Bei jedem *T*-Test werden alle Fälle verwendet, die gültige Daten für die getestete Variable aufweisen. Die Stichprobengröße kann von Test zu Test unterschiedlich ausfallen.
- **Listenweiser Fallausschluß.** Jeder *T*-Test verwendet nur Fälle, die gültige Daten für alle Variablen aufweisen, die in einem der angeforderten *T*-Tests verwendet werden. Die Stichprobengröße bleibt bei allen Tests konstant.

Zusätzliche Funktionen beim Befehl T-TEST

Mit der SPSS-Befehlssprache verfügen Sie über die folgenden zusätzlichen Möglichkeiten:

- Erstellen von T-Tests für eine Stichprobe sowie für unabhängige Stichproben mit einem einzigen Befehl.
- Testen einer Variablen gegen alle Variablen in einer Liste mit einem gepaarten T-Test (mit dem Unterbefehl `PAIRS`).

Vollständige Informationen zur Syntax finden Sie in der *SPSS Command Syntax Reference*.

Kapitel
22

Einfaktorielle ANOVA

Die Prozedur Einfaktorielle ANOVA führt eine einfaktorielle Varianzanalyse für eine quantitative abhängige Variable mit einer einzelnen (unabhängigen) Faktorvariablen durch. Mit der Varianzanalyse wird die Hypothese überprüft, daß mehrere Mittelwerte gleich sind. Dieses Verfahren ist eine Erweiterung des *T*-Tests bei zwei Stichproben.

Sie können zusätzlich zur Feststellung, daß Differenzen zwischen Mittelwerten vorhanden sind, auch bestimmen, welche Mittelwerte abweichen. Für den Vergleich von Mittelwerten gibt es zwei Arten von Tests: A-priori-Kontraste und Post-Hoc-Tests. Kontraste sind Tests, die *vor* der Ausführung des Experiments eingerichtet werden, Post-Hoc-Tests werden *nach* dem Experiment ausgeführt. Sie können auch auf Trends für mehrere Kategorien testen.

Beispiel. Paniertes Fleisch absorbiert beim Fritieren unterschiedliche Mengen an Fett. Ein Experiment wird mit den folgenden drei Fettsorten durchgeführt: Distelöl, Maiskeimöl und Schmalz. Distelöl und Maiskeimöl sind ungesättigte Fette, Schmalz ist ein gesättigtes Fett. Sie können bestimmen, ob die Menge des absorbierten Fetts von der Fettsorte abhängt. Gleichzeitig können Sie einen A-priori-Kontrast einrichten, um zu ermitteln, ob sich die absorbierte Fettmenge bei gesättigten und ungesättigten Fetten unterscheidet.

Statistiken. Für jede Gruppe: Anzahl der Fälle, Mittelwert, Standardabweichung, Standardfehler des Mittelwerts, Minimum, Maximum und 95 %-Konfidenzintervall für den Mittelwert. Levene-Test auf Homogenität der Varianzen, Varianzanalyse-Tabellen und zuverlässige Tests auf Gleichheit der Mittelwerte für jede abhängige Variable, benutzerspezifische A-priori-Kontraste, Post-Hoc-Spannweitentests und Mehrfachvergleiche: Bonferroni, Sidak, ehrlich signifikante Differenz nach Tukey, GT2 nach Hochberg, Gabriel, Dunnett, Ryan-Einot-Gabriel-Welsch *F*-Test (*F* nach R-E-G-W), Spannweitentest nach Ryan-Einot-Gabriel-Welsch (*Q* nach R-E-G-W), Tamhane-T2, Dunnett-T3, Games-Howell, Dunnett-*C*, Duncans multipler Spannweitentest, Student-Newman-Keuls (S-N-K), Tukey-*b*, Waller-Duncan, Scheffé,und geringste signifikante Differenz.

Daten. Die Werte der Faktorvariablen müssen ganzzahlig sein, die abhängige Variable muß quantitativ sein (Messung auf Intervallebene).

Annahmen. Jede Gruppe bildet eine unabhängige zufällige Stichprobe aus einer normalverteilten Grundgesamtheit. Die Varianzanalyse ist unempfindlich gegenüber Abweichungen von der Normalverteilung. Die Daten müssen jedoch symmetrisch verteilt sein. Die Gruppen müssen aus Grundgesamtheiten mit gleichen Varianzen stammen. Sie überprüfen diese Annahme mit Hilfe des Levene-Tests auf Homogenität der Varianzen.

Abbildung 22-1
Ausgabe der einfaktoriellen ANOVA

Deskriptive Statistik

		N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
						Untergrenze	Obergrenze		
FETT	Erdnuß-Öl	6	72.00	13.34	5.45	58.00	86.00	56	95
	Schweineschmalz	6	85.00	7.77	3.17	76.84	93.16	77	97
	Mais-Öl	6	62.00	8.22	3.36	53.37	70.63	49	70
	Gesamt	18	73.00	13.56	3.20	66.26	79.74	49	97

Test der Homogenität der Varianzen

	Levene-Statistik	df1	df2	Signifikanz
FETT	.534	2	15	.597

ANOVA

		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
FETT	Zwischen den Gruppen	1596.000	2	798.000	7.824	.005
	Innerhalb der Gruppen	1530.000	15	102.000		
	Gesamt	3126.000	17			

Kontrast-Koeffizienten

		ÖL		
		Erdnuß-Öl	Schweineschmalz	Mais-Öl
Kontrast	1	-.5	1	-.5

Kontrast-Tests

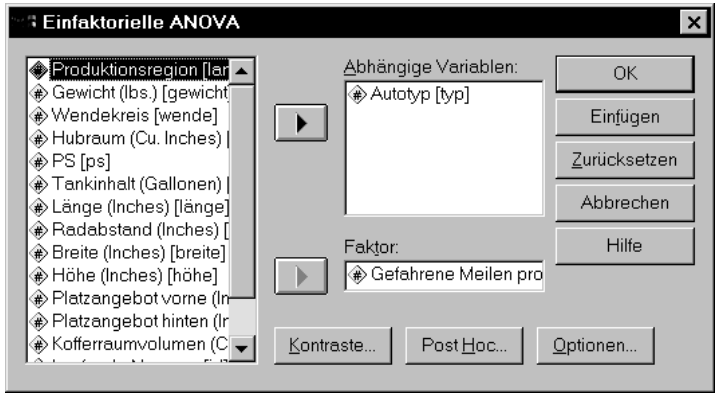
				Kontrastwert	Standardfehler	T	df	Signifikanz (2-seitig)
FETT	Varianzen sind gleich	Kontrast	1	18.00	5.05	3.565	15	.003
	Varianzen sind nicht gleich	Kontrast	1	18.00	4.51	3.995	12.542	.002

So lassen Sie eine einfaktorielle ANOVA berechnen

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Mittelwerte vergleichen
    Einfaktorielle ANOVA…

Abbildung 22-2
Dialogfeld "Einfaktorielle ANOVA"

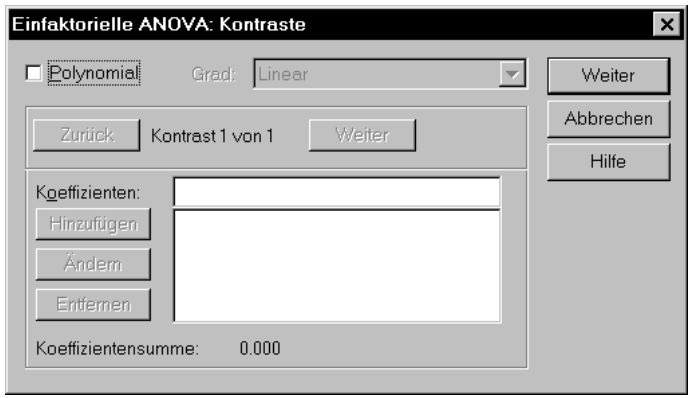


▶ Wählen Sie eine oder mehrere abhängige Variablen aus.

▶ Wählen Sie eine unabhängige Faktorvariable aus.

Einfaktorielle ANOVA: Kontraste

Abbildung 22-3
Dialogfeld "Einfaktorielle ANOVA: Kontraste"



Sie können die Quadratsummen zwischen den Gruppen in Trend-Komponenten zerlegen oder A-priori-Kontraste festlegen.

Polynomial. Damit zerlegen Sie die Quadratsummen zwischen den Gruppen in Trend-Komponenten. Sie können die abhängige Variable auf einen Trend über die geordneten Stufen der Faktorvariablen prüfen. Sie können beispielsweise prüfen, ob beim Gehalt über die geordneten Stufen des höchsten erreichten akademischen Grads ein linearer (steigender oder fallender) Trend vorliegt.

- **Grad.** Sie können Polynome ersten, zweiten, dritten, vierten und fünften Grades auswählen.

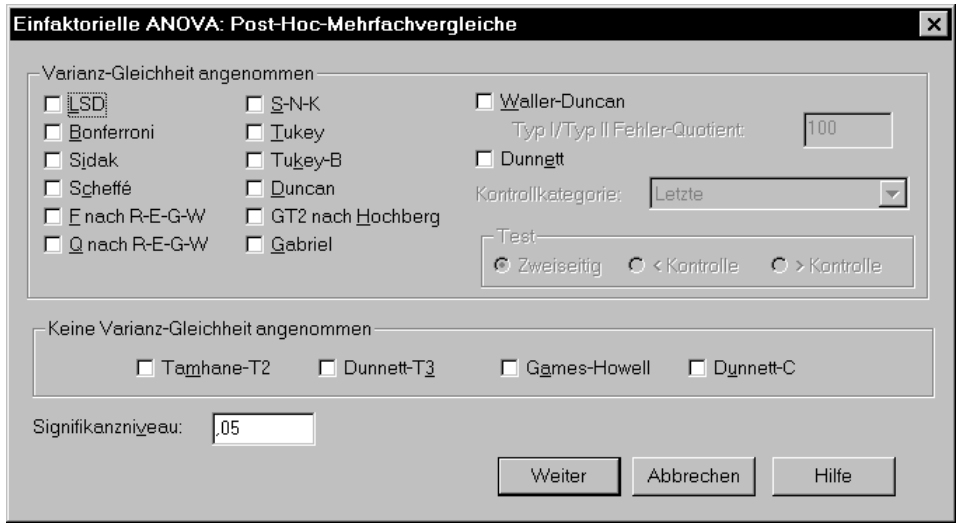
Koeffizienten. Mit der *T*-Statistik werden benutzerdefinierte A-priori-Kontraste getestet. Geben Sie für jede Gruppe (Kategorie) der Faktorvariablen einen Koeffizienten ein, und klicken Sie nach jeder Eingabe auf Hinzufügen. Jeder neue Wert wird am Ende der Liste der Koeffizienten hinzugefügt. Um zusätzliche Kontrastgruppen festzulegen, klicken Sie auf Weiter. Verwenden Sie Weiter und Zurück, um zwischen den Kontrastgruppen zu wechseln.

Die Reihenfolge der Koeffizienten ist wichtig, weil sie den aufsteigend geordneten Kategoriewerten der Faktorvariablen entspricht. Der erste Koeffizient der Liste entspricht dem kleinsten Gruppenwert der Faktorvariablen, der letzte Koeffizient dem größten Wert. Bei zum Beispiel sechs Kategorien der Faktorvariablen stellen die Koeffizienten –1, 0, 0, 0,5 und 0,5 einen Kontrast zwischen der ersten und der fünften und sechsten Gruppe her. Bei den meisten Anwendungen muß die Summe der

Koeffizienten 0 ergeben. Sie können auch Werte benutzen, deren Summe ungleich 0 ist. In diesem Fall wird jedoch eine Warnung angezeigt.

Einfaktorielle ANOVA: Post-Hoc-Mehrfachvergleiche

Abbildung 22-4
Dialogfeld "Einfaktorielle ANOVA: Post-Hoc-Mehrfachvergleiche"



Sobald Sie festgestellt haben, daß es Abweichungen zwischen den Mittelwerten gibt, können Sie mit Post-Hoc-Spannweiten-Tests und paarweisen multiplen Vergleichen untersuchen, welche Mittelwerte sich unterscheiden. Spannweitentests ermitteln homogene Untergruppen von Mittelwerten, die nicht voneinander abweichen. Mit paarweisen Mehrfachvergleichen testen Sie die Differenz zwischen gepaarten Mittelwerten. Die Ergebnisse werden in einer Matrix angezeigt, in der Gruppenmittelwerte, die auf einem Alpha-Niveau von 0,05 signifikant voneinander abweichen, durch Sterne markiert sind.

Varianz-Gleichheit angenommen

Die ehrlich signifikante Differenz nach Tukey, der GT2 nach Hochberg, der Gabriel-Test und der Scheffé-Test sind Tests für Mehrfachvergleiche und Spannweitentests. Andere Spannweitentests sind Tukey-*B*, S-N-K

(Student-Newman-Keuls), Duncan, *F* nach R-E-G-W (*F*-Test nach Ryan-Einot-Gabriel-Welsch), *Q* nach R-E-G-W (Spannweitentest nach Ryan-Einot-Gabriel-Welsch) und Waller-Duncan. Verfügbare Tests für Mehrfachvergleiche sind Bonferroni, ehrlich signifikante Differenz nach Tukey, Sidak, Gabriel, Hochberg, Dunnett, Scheffé und LSD (geringste signifikante Differenz).

- **LSD.** T-Tests werden durchgeführt, um alle paarweisen Vergleiche von Gruppenmittelwerten durchzuführen. Es erfolgt keine Korrektur der Fehlerrate bei Mehrfachvergleichen.
- **Bonferroni.** Führt paarweise Vergleiche zwischen Gruppenmittelwerten mit T-Tests aus; regelt dabei jedoch auch die Gesamtfehlerrate, indem die Fehlerrate für jeden Test auf den Quotienten aus der experimentellen Fehlerrate und der Gesamtzahl der Tests gesetzt wird. Dadurch wird das beobachtete Signifikanzniveau an Mehrfachvergleiche angepaßt.
- **Sidak.** Ein paarweiser multipler Vergleichstest, basierend auf einer T-Statistik. Bei Sidaks Test wird das Signifikanzniveau für die multiplen Vergleiche korrigiert und engere Grenzen vergeben als bei Bonferroni.
- **Scheffé-Prozedur.** Führt gemeinsame paarweise Vergleiche gleichzeitig für alle möglichen paarweisen Kombinationen der Mittelwerte durch. Verwendet die F-Stichprobenverteilung. Dieser Test kann verwendet werden, um nicht nur paarweise Vergleiche durchzuführen, sondern alle möglichen linearen Kombinationen von Gruppenmittelwerten zu untersuchen.
- **F nach R-E-G-W.** Mehrfaches Rückschrittverfahren nach Ryan-Einot-Gabriel-Welsh, basierend auf einem F-Test.
- **Q nach R-E-G-W.** Mehrfaches Rückschrittverfahren nach Ryan-Einot-Gabriel-Welsh, das auf der studentisierten Spannweite beruht.
- **S-N-K.** Führt alle paarweisen Vergleiche zwischen Mittelwerten unter Verwendung der t-Verteilung aus. Bei gleich großen Stichproben werden auch die Mittelwertpaare innerhalb homogener Untergruppen verglichen; dabei wird ein schrittweises Verfahren verwendet. Die Mittelwerte werden vom größten zum kleinsten Wert sortiert, extreme Differenzen werden zuerst getestet.
- **Tukey.** Verwendet die Student-Verteilung für alle möglichen paarweisen Vergleiche zwischen den Gruppen. Setzt die Fehlerrate für das Experiment gleich der Fehlerrate für die Gesamtheit aller paarweisen Vergleiche.

- **Tukey-B-Test.** Verwendet die Student-Verteilung für paarweise Vergleiche zwischen Gruppen. Der kritische Wert ist der Durchschnitt des entsprechenden Werts für die ehrlich signifikante Differenz nach Tukey und für Student-Newman-Keuls.
- **Duncan.** Bei diesem Test werden paarweise Vergleiche angestellt, deren schrittweise Reihenfolge identisch ist mit der Reihenfolge, die beim Student-Newman-Keuls-Test verwendet wird. Abweichend wird aber ein Sicherheitsniveau für die Fehlerrate der zusammengefaßten Tests statt einer Fehlerrate für die einzelnen Tests gesetzt. Es wird die studentisierte Spannweiten-Statistik verwendet.
- **GT2 nach Hochberg.** Ein multipler Vergleichs- und Spannweitentest, der das studentisierte Maximalmodul verwendet. Ähnelt dem Test auf ehrlich signifikante Differenz nach Tukey.
- **Gabriel.** Ein paarweiser Vergleichstest, der das studentisierte Maximalmodul verwendet. Er ist in der Regel aussagekräftiger als der GT2-Test nach Hochberg, wenn die Zellengrößen ungleich sind. Der Test nach Gabriel kann ungenau sein, wenn die Zellengrößen große Abweichungen aufweisen.
- **Waller-Duncan.** Ein Test für Mehrfachvergleiche auf der Grundlage einer T-Statistik; verwendet eine Bayessche Methode.
- **Dunnett.** Ein T-Test mit paarweisen Mehrfachvergleichen, der ein Set von Verarbeitungen mit einem einzelnen Kontrollmittelwert vergleicht. Als Kontrollkategorie ist die letzte Kategorie voreingestellt. Sie können aber auch die erste Kategorie einstellen. Verwenden Sie einen zweiseitigen Test, um zu überprüfen, ob sich der Mittelwert bei jeder Stufe (außer der Kontrollkategorie) des Faktors von dem Mittelwert der Kontrollkategorie unterscheidet. Wählen Sie <>Kontrolle, um zu überprüfen, ob der Mittelwert bei allen Stufen des Faktors kleiner als der Mittelwert der Kontrollkategorie ist. Wählen Sie >Kontrolle, um zu überprüfen, ob der Mittelwert bei allen Stufen des Faktors größer als der Mittelwert der Kontrollkategorie ist.

Keine Varianz-Gleichheit angenommen

Tests für Mehrfachvergleiche, die keine Varianzgleichheit voraussetzen, sind Tamhane-T2, Dunnett-T3, Games-Howell und Dunnett-*C*.

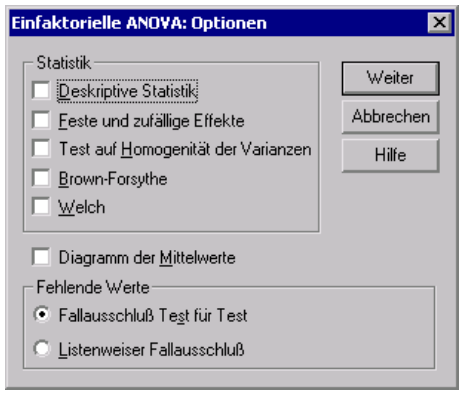
- **T2 nach Tamhane.** Konservative, paarweise Vergleichstests auf der Grundlage eines T-Tests. Dieser Test ist für ungleiche Varianzen geeignet sind.

- **T3 nach Dunnett.** Ein Test für paarweise Vergleiche, der auf dem studentisierten maximalen Modalwert beruht. Dieser Test ist geeignet, wenn die Varianzen verschieden sind.
- **Games-Howell.** Ein manchmal schwacher, paarweiser Vergleichstest. Dieser Test ist geeignet, wenn die Varianzen ungleich sind.
- **C nach Dunnett.** Ein paarweiser Vergleichstest, basierend auf dem studentisierten Bereich. Dieser Test ist geeignet, wenn die Varianzen ungleich sind.

Anmerkung: Die Ausgabe von Post-Hoc-Tests läßt sich oft einfacher interpretieren, wenn Sie im Dialogfeld "Tabelleneigenschaften" die Option Leere Zeilen und Spalten ausblenden deaktivieren. (In einer aktivierten Pivot-Tabelle: Tabelleneigenschaften im Menü "Format".)

Einfaktorielle ANOVA: Optionen

Abbildung 22-5
Dialogfeld "Einfaktorielle ANOVA: Optionen"



Statistiken. Wählen Sie mindestens eine der folgenden Optionen aus:

- **Deskriptive Statistik.** Hiermit berechnen Sie Anzahl der Fälle, Mittelwert, Standardabweichung, Standardfehler des Mittelwerts, Minimum, Maximum und das 95 %-Konfidenzintervall für jede abhängige Variable in jeder Gruppe.
- **Feste und zufällige Effekte.** Hiermit werden die Standardabweichung, der Standardfehler und das 95 %-Konfidenzintervall für das Modell mit festen Effekten sowie der Standardfehler, das 95%-Konfidenzintervall und die

Schätzung der Varianz zwischen Komponenten für das Modell mit zufälligen Effekten angezeigt.

- **Test auf Homogenität der Varianzen.** Bei dieser Option wird die Levene-Statistik berechnet, mit der Sie die Gruppenvarianzen auf Gleichheit testen können. Dieser Test setzt keine Normalverteilung voraus.
- **Brown-Forsythe.** Bei dieser Option wird die Brown-Forsythe-Statistik berechnet, mit der Sie die Gruppenmittelwerte auf Gleichheit testen können. Diese Statistik ist der *F*-Statistik vorzuziehen, wenn die Annahme gleicher Varianzen sich nicht bestätigt.
- **Welch.** Bei dieser Option wird die Welch-Statistik berechnet, mit der Sie die Gruppenmittelwerte auf Gleichheit testen können. Diese Statistik ist der *F*-Statistik vorzuziehen, wenn die Annahme gleicher Varianzen sich nicht bestätigt.

Diagramm der Mittelwerte. Bei dieser Option wird ein Diagramm für die Mittelwerte der Untergruppen ausgegeben. Dabei handelt es sich um die Mittelwerte für jede Gruppe, die durch die Werte der Faktorvariablen definiert ist.

Fehlende Werte. Bestimmt die Verarbeitung fehlender Werte.

- **Fallausschluß Test für Test.** Bei Auswahl dieser Option werden Fälle mit einem fehlenden Wert für die abhängige Variable oder die Faktorvariable in einer bestimmten Analyse in dieser Analyse nicht verwendet. Ein Fall wird außerdem nicht verwendet, wenn er außerhalb des Bereichs liegt, der für die Faktorvariable definiert ist.
- **Listenweiser Fallausschluß.** Fälle mit fehlenden Werten für die Faktorvariable oder eine abhängige Variable, die in der Liste der abhängigen Variablen des Hauptdialogfelds enthalten sind, werden aus allen Analysen ausgeschlossen. Wenn Sie nicht mehrere abhängige Variablen festgelegt haben, hat dies keine Auswirkung.

Zusätzliche Funktionen beim Befehl ONEWAY

Mit der SPSS-Befehlssprache verfügen Sie über die folgenden zusätzlichen Möglichkeiten:

- Erstellen von Statistiken mit festen und zufälligen Effekten. Standardabweichung, Standardfehler des Mittelwerts und 95%-Konfidenzintervalle für ein Modell mit festen Effekten. Standardfehler, 95%-Konfidenzintervalle und die Schätzung der

Varianz zwischen Komponenten für ein Modell mit zufälligen Effekten (mit STATISTICS=EFFECTS).

- Angeben der Alpha-Niveaus für die Test für Mehrfachvergleiche auf geringste signifikante Differenz sowie nach Bonferroni, Duncan und Scheff (mit dem Unterbefehl RANGES).
- Schreiben einer Matrix der Mittelwerte, Standardabweichungen und Häufigkeiten, oder Lesen einer Matrix der Mittelwerte, Häufigkeiten, gemeinsame Varianzen sowie der Freiheitsgrade für die gemeinsamen Varianzen. Diese Matrizen können anstellen der Rohdaten verwendet werden, um eine einfaktorielle Analyse der Varianz durchzuführen (mit dem Unterbefehl MATRIX).

Vollständige Informationen zur Syntax finden Sie in der *SPSS Command Syntax Reference*.

Kapitel
23

GLM - Univariat

Mit der Prozedur “GLM - Univariat” können Sie Regressionsanalysen und Varianzanalysen für eine abhängige Variable mit einem oder mehreren Faktoren und/oder Variablen durchführen. Die Faktorvariablen unterteilen die Grundgesamtheit in Gruppen. Unter Verwendung dieser auf einem allgemeinen linearen Modell basierenden Prozedur können Sie Nullhypothesen über die Effekte anderer Variablen auf die Mittelwerte verschiedener Gruppierungen einer einzelnen abhängigen Variablen testen. Sie können die Wechselwirkungen zwischen Faktoren und die Effekte einzelner Faktoren untersuchen, von denen einige zufällig sein können. Außerdem können Sie die Auswirkungen von Kovariaten und Wechselwirkungen zwischen Kovariaten und Faktoren berücksichtigen. Bei der Regressionsanalyse werden die unabhängigen Variablen (Einflußvariablen) als Kovariaten angegeben.

Es können sowohl ausgeglichene als auch nicht ausgeglichene Modelle getestet werden. Ein Design ist ausgeglichen, wenn jede Zelle im Modell dieselbe Anzahl von Fällen enthält. Mit der Prozedur “GLM - Univariat” werden nicht nur Hypothesen getestet, sondern zugleich Parameter geschätzt.

Zum Testen von Hypothesen stehen häufig verwendete a-priori-Kontraste zur Verfügung. Nachdem die Signifikanz mit einem *F*-Gesamttest nachgewiesen wurde, können Sie Post-Hoc-Tests verwenden, um Differenzen zwischen bestimmten Mittelwerten berechnen zu lassen. Geschätzte Randmittel dienen als Schätzer für die vorhergesagten Mittelwerte der Zellen im Modell, und mit Profilplots (Wechselwirkungsdiagrammen) dieser Mittelwerte können Sie einige dieser Beziehungen in einfacher Weise visuell darstellen.

Residuen, Einflußwerte, die Cook-Distanz und Hebelwerte können zum Überprüfen von Annahmen als neue Variablen in der Datendatei gespeichert werden.

Mit der WLS-Gewichtung können Sie eine Variable angeben, um Beobachtungen für eine WLS-Analyse (Weighted Least Squares, deutsch: gewichtete kleinste Quadrate) unterschiedlich zu gewichten. Dies kann notwendig sein, um etwaige Unterschiede in der Präzision von Messungen auszugleichen.

Beispiel. Im Rahmen einer sportwissenschaftlichen Studie beim Berlin-Marathon werden mehrere Jahre lang Daten über einzelne Läufer aufgenommen. Die abhängige Variable ist die Zeit, die jeder Läufer für die Strecke benötigt. Andere berücksichtigte Faktoren sind beispielsweise das Wetter (kalt, angenehm oder heiß), die Anzahl von Trainingsmonaten, die Anzahl der bereits absolvierten Marathons und das Geschlecht. Das Alter der betreffenden Personen wird als Kovariate betrachtet. Ein mögliches Ergebnis wäre, daß das Geschlecht ein signifikanter Effekt und die Wechselwirkung von Geschlecht und Wetter signifikant ist.

Methoden. Zum Überprüfen der verschiedenen Hypothesen können Quadratsummen vom Typ I, Typ II, Typ III und Typ IV verwendet werden. Die Voreinstellung sieht den Typ III vor.

Statistik. Post-Hoc-Spannweitentests und Mehrfachvergleiche: geringste signifikante Differenz, Bonferroni, Sidak, Scheffé, multiples *F* nach Ryan-Einot-Gabriel-Welsch, multiple Spannweite nach Ryan-Einot-Gabriel-Welsch, Student-Newman-Keuls-Test, ehrlich signifikante Differenz nach Tukey, Tukey-*B*, Duncan, GT2 nach Hochberg, Gabriel, Waller-Duncan-*T*-Test, Dunnett (einseitig und zweiseitig), Tamhane-T2, Dunnett-T3, Games-Howell und Dunnett-*C*. Deskriptive Statistiken: beobachtete Mittelwerte, Standardabweichungen und Häufigkeiten aller abhängigen Variablen in allen Zellen. Levene-Test auf Homogenität der Varianzen.

Diagramme. Streubreite gegen mittleres Niveau, Residuen-Diagramme, Profilplots (Wechselwirkung).

Daten. Die abhängige Variable ist quantitativ. Faktoren sind kategorial. Sie können numerische Werte oder String-Werte von bis zu acht Zeichen Länge annehmen. Kovariaten sind quantitative Variablen, die mit der abhängigen Variablen in Beziehung stehen.

Annahmen. Die Daten sind eine Stichprobe aus einer normalverteilten Grundgesamtheit. In der Grundgesamtheit sind alle Zellenvarianzen gleich. Die Varianzanalyse ist unempfindlich gegenüber Abweichungen von der Normalverteilung. Die Daten müssen jedoch symmetrisch verteilt sein. Zum Überprüfen der Annahmen können Sie Tests auf Homogenität der Varianzen vornehmen und Diagramme der Streubreite gegen das mittlere Niveau ausgeben lassen. Sie können auch die Residuen untersuchen und Residuen-Diagramme anzeigen lassen.

Abbildung 23-1
Ausgabe von GLM - Univariat

Tests of Between-Subjects Effects

Dependent Variable: SPVOL

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	22.520[1]	11	2.047	12.376	.000
Intercept	1016.981	1	1016.981	6147.938	.000
Flour	8.691	3	2.897	17.513	.000
Fat	10.118	2	5.059	30.583	.000
Surfactant	.997	2	.499	3.014	.082
Fat*Surfactant	5.639	4	1.410	8.522	.001
Error	2.316	14	.165		
Total	1112.960	26			
Corrected Total	24.835	25			

1. R Squared = .907 (Adjusted R Squared = .833)

fat * surfactant

Dependent Variable: SPVOL

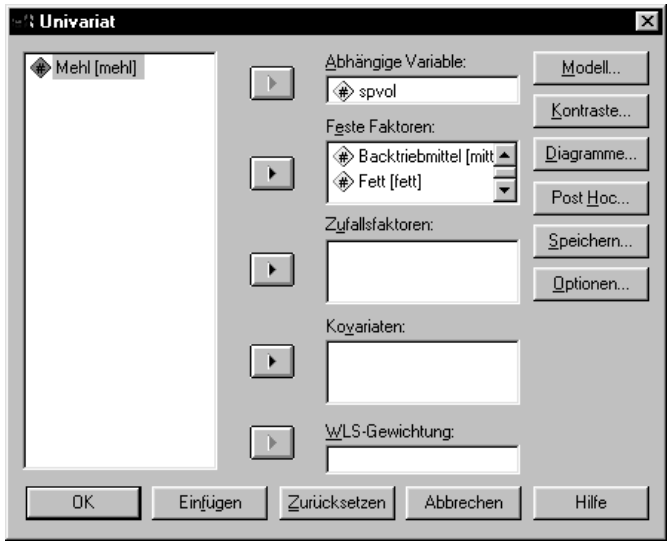
fat	surfactant	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1	5.536	.240	5.021	6.052
	2	5.891	.239	5.378	6.404
	3	6.123	.241	5.605	6.641
2	1	7.023	.241	6.505	7.541
	2	6.708	.301	6.064	7.353
	3	6.000	.203	5.564	6.436
3	1	6.629	.301	5.984	7.274
	2	7.200	.203	6.764	7.636
	3	8.589	.300	7.945	9.233

So berechnen Sie eine univariate Analyse der Varianz (GLM):

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Allgemeines lineares Modell
    Univariat...

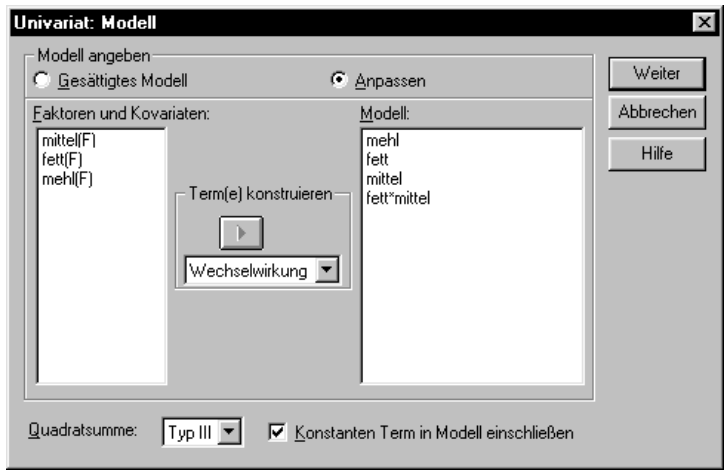
Abbildung 23-2
Dialogfeld "Univariat"



- Wählen Sie eine abhängige Variable aus.
- Wählen Sie in Abhängigkeit von den Daten Variablen als feste Faktoren, Zufallsfaktoren und Kovariaten aus.
- Optional können Sie mit der WLS-Gewichtung eine Gewichtungsvariable für WLS-Analyse (Weighted Least Squares, gewichtete kleinste Quadrate) angeben. Wenn der Wert der Gewichtungsvariablen null, negativ oder fehlend ist, wird der Fall aus der Analyse ausgeschlossen. Eine bereits im Model verwendete Variable kann nicht als Gewichtungsvariable verwendet werden.

GLM: Modell

Abbildung 23-3
Dialogfeld "Univariat: Model"



Modell angeben. Ein gesättigtes Modell enthält alle Faktoren-Haupteffekte, alle Kovariaten-Haupteffekte und alle faktorweisen Wechselwirkungen. Es enthält keine Kovariaten-Wechselwirkungen. Wählen Sie Anpassen aus, um nur eine Teilmenge von Wechselwirkungen oder Wechselwirkungen zwischen Faktoren und Kovariaten festzulegen. Sie müssen alle in das Modell zu übernehmenden Terme angeben.

Faktoren und Kovariaten. Die Faktoren und Kovariaten werden in der Liste mit einem (F) für "Fester Faktor" und einem (C) für "Kovariate" gekennzeichnet. Bei der univariaten Analyse kennzeichnet (R) einen Zufallsfaktor.

Modell. Das Modell ist von der Art Ihrer Daten abhängig. Nach der Auswahl von Anpassen können Sie die Haupteffekte und Wechselwirkungen auswählen, die für Ihre Analyse von Interesse sind.

Quadratsumme. Hier wird die Methode zum Berechnen der Quadratsumme festgelegt. Für ausgeglichene und unausgeglichene Modelle ohne fehlende Zellen wird meistens die Methode mit Quadratsummen vom Typ III angewendet.

Konstanten Term in Modell einschließen. Der konstante Term wird gewöhnlich in das Modell aufgenommen. Falls Sie sicher sind, daß die Daten durch den Koordinatenursprung verlaufen, können Sie den konstanten Term ausschließen.

Terme konstruieren

Für die ausgewählten Faktoren und Kovariaten:

Wechselwirkung. Hiermit wird der Wechselwirkungsterm mit der höchsten Ordnung von allen ausgewählten Variablen erzeugt. Dies ist die Standardeinstellung.

Haupteffekte. Legt einen Haupteffekt-Term für jede ausgewählte Variable an.

Alle 2-Weg. Hiermit werden alle möglichen 2-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 3-Weg. Hiermit werden alle möglichen 3-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 4-Weg. Hiermit werden alle möglichen 4-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 5-Weg. Hiermit werden alle möglichen 5-Weg-Wechselwirkungen der ausgewählten Variablen erzeugt.

Quadratsumme

Für das Modell können Sie einen Typ von Quadratsumme auswählen. Typ III wird am häufigsten verwendet und ist die Standardeinstellung.

Typ I. Diese Methode ist auch als die Methode der hierarchischen Zerlegung der Quadratsummen bekannt. Jeder Term wird nur für den Vorläuferterm im Modell angepaßt. Quadratsummen vom Typ I werden gewöhnlich in den folgenden Situationen verwendet:

- Ein ausgeglichenes ANOVA-Modell, in dem alle Haupteffekte vor den Wechselwirkungseffekten 1. Ordnung festgelegt werden, alle Wechselwirkungseffekte 1. Ordnung wiederum vor den Wechselwirkungseffekten 2. Ordnung festgelegt werden und so weiter.
- Ein polynomiales Regressionsmodell, in dem alle Terme niedrigerer Ordnung vor den Termen höherer Ordnung festgelegt werden.
- Ein rein verschachteltes Modell, in welchem der zuerst bestimmte Effekt in dem als zweiten bestimmten Effekt verschachtelt ist, der zweite Effekt wiederum im dritten und so weiter. (Diese Form der Verschachtelung kann nur durch Verwendung der Befehlssprache erreicht werden.)

Typ II. Bei dieser Methode wird die Quadratsumme eines Effekts im Modell angepaßt an alle anderen "zutreffenden" Effekte berechnet. Ein zutreffender Effekt ist ein Effekt, der mit allen Effekten in Beziehung steht, die den untersuchten Effekt nicht enthalten. Die Methode mit Quadratsummen vom Typ II wird gewöhnlich in den folgenden Fällen verwendet:

- Bei ausgeglichenen ANOVA-Modellen.
- Bei Modellen, die nur Haupteffekte von Faktoren enthalten.
- Bei Regressionsmodellen.
- Bei rein verschachtelten Designs. (Diese Form der Verschachtelung kann durch Verwendung der Befehlssprache erreicht werden.)

Typ III. Voreinstellung. Bei dieser Methode werden die Quadratsummen eines Effekts im Design als Quadratsummen orthogonal zu den Effekten (sofern vorhanden), die den Effekt enthalten, und mit Bereinigung um alle anderen Effekte, die diesen Effekt nicht enthalten, berechnet. Der große Vorteil der Quadratsummen vom Typ III ist, daß sie invariant bezüglich der Zellenhäufigkeiten sind, solange die allgemeine Form der Schätzbarkeit konstant bleibt. Daher wird dieser Typ von Quadratsumme oft für nicht ausgeglichene Modelle ohne fehlende Zellen als geeignet angesehen. In einem faktoriellen Design ohne fehlende Zellen ist diese Methode äquivalent zu der Methode der gewichteten Mittelwertquadrate nach Yates. Die Methode mit Quadratsummen vom Typ III wird gewöhnlich in folgenden Fällen verwendet:

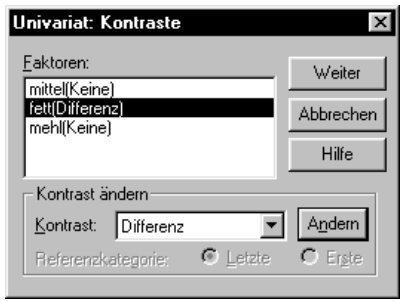
- Alle bei Typ I und Typ II aufgeführten Modelle.
- Alle ausgeglichenen oder unausgeglichenen Modelle ohne leere Zellen.

Typ IV. Diese Methode ist dann geeignet, wenn es keine fehlenden Zellen gibt. Für alle Effekte *F* im Design: Wenn *F* in keinem anderen Effekt enthalten ist, dann gilt: Typ IV = Typ III = Typ II. Wenn *F* in anderen Effekten enthalten ist, werden bei Typ IV die Kontraste zwischen den Parametern in *F* gleichmäßig auf alle Effekte höherer Ordnung verteilt. Die Methode mit Quadratsummen vom Typ IV wird gewöhnlich in folgenden Fällen verwendet:

- Alle bei Typ I und Typ II aufgeführten Modelle.
- Alle ausgeglichenen oder unausgeglichenen Modelle mit leeren Zellen.

GLM: Kontraste

Abbildung 23-4
Dialogfeld "Univariat: Kontraste"



Kontraste werden verwendet, um auf Unterschiede zwischen den Stufen eines Faktors zu testen. Für jeden Faktor im Modell kann ein Kontrast festgelegt werden (in einem Modell mit Meßwiederholungen für jeden Zwischensubjektfaktor). Kontraste stellen lineare Kombinationen der Parameter dar.

Das Testen der Hypothesen basiert auf der Nullhypothese **LB** = 0. Dabei ist **L** die Kontrastkoeffizienten-Matrix und **B** der Parametervektor. Wenn Sie einen Kontrast festlegen, legt SPSS eine **L**-Matrix an, bei der die Spalten für den betreffenden Faktor mit dem Kontrast übereinstimmen. Die verbleibenden Spalten werden so angepaßt, daß die **L**-Matrix schätzbar ist.

Die Ausgabe beinhaltet eine *F*-Statistik für jedes Set von Kontrasten. Für die Kontrastdifferenzen werden außerdem simultane Konfidenzintervalle nach Bonferroni auf der Grundlage der Student-*T*-Verteilung angezeigt.

Verfügbare Kontraste

Als Kontraste sind "Abweichung", "Einfach", "Differenz", "Helmert", "Wiederholt" und "Polynomial" verfügbar. Bei Abweichungskontrasten und einfachen Kontrasten können Sie wählen, ob die letzte oder die erste Kategorie als Referenzkategorie dient.

Kontrasttypen

Abweichung. Vergleicht den Mittelwert jeder Faktorstufe (außer bei Referenzkategorien) mit dem Mittelwert aller Faktorstufen (Gesamtmittelwert). Die Stufen des Faktors können in beliebiger Ordnung vorliegen.

Einfach. Vergleicht den Mittelwert jeder Faktorstufe mit dem Mittelwert einer angegebenen Faktorstufe. Dieser Kontrasttyp ist nützlich, wenn es eine Kontrollgruppe gibt. Sie können die erste oder die letzte Kategorie als Referenz auswählen.

Differenz. Vergleicht den Mittelwert jeder Faktorstufe (außer der ersten) mit dem Mittelwert der vorhergehenden Faktorstufen. (Dies wird gelegentlich auch als umgekehrter Helmert-Kontrast bezeichnet).

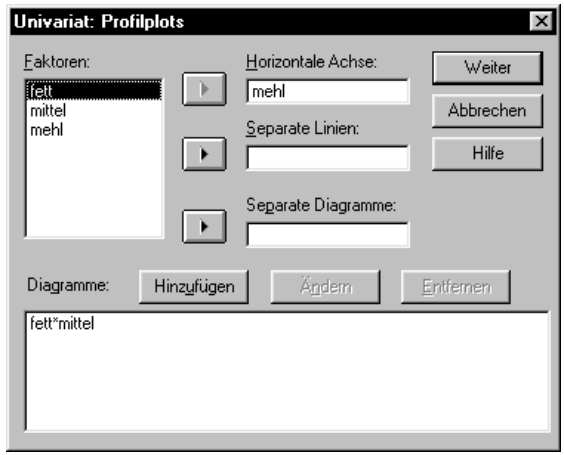
Helmert. Vergleicht den Mittelwert jeder Stufe des Faktors (bis auf die letzte) mit dem Mittelwert der folgenden Stufen.

Wiederholt. Vergleicht den Mittelwert jeder Faktorstufe (außer der letzten) mit dem Mittelwert der folgenden Faktorstufe.

Polynomial. Vergleicht den linearen Effekt, quadratischen Effekt, kubischen Effekt und so weiter. Der erste Freiheitsgrad enthält den linearen Effekt über alle Kategorien; der zweite Freiheitsgrad den quadratischen Effekt und so weiter. Die Kontraste werden oft verwendet, um polynomiale Trends zu schätzen.

GLM: Profilplots

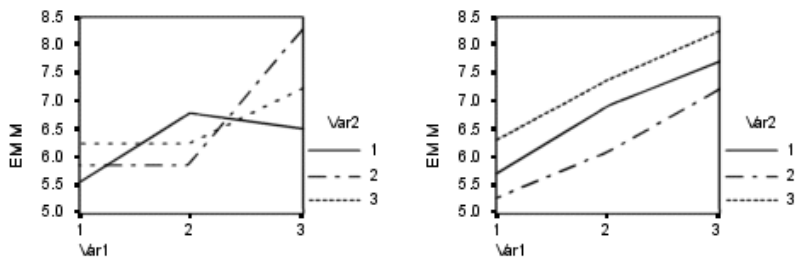
Abbildung 23-5
Dialogfeld "Univariat: Profilplots"



Profilplots (Wechselwirkungsdiagramme) sind hilfreich zum Vergleichen von Randmitteln im Modell. Ein Profilplot ist ein Liniendiagramm, in dem jeder Punkt das geschätzte Randmittel einer abhängigen Variablen (angepaßt an die Kovariaten) bei einer Stufe eines Faktors angibt. Die Stufen eines zweiten Faktors können zum Erzeugen getrennter Linien verwendet werden. Jede Stufe in einem dritten Faktor kann verwendet werden, um ein separates Diagramm zu erzeugen. Alle festen Faktoren und Zufallsfaktoren (sofern vorhanden) sind für Diagramme verfügbar. Bei multivariaten Analysen werden Profilplots für jede abhängige Variable erstellt. Bei einer Analyse mit Meßwiederholungen können in Profilplots sowohl Zwischensubjektfaktoren als auch Innersubjektfaktoren verwendet werden. "GLM – Multivariat" und "GLM – Meßwiederholungen" sind nur verfügbar, wenn Sie SPSS Advanced Models installiert haben.

Ein Profilplot für einen Faktor zeigt, ob die geschätzten Randmittel mit den Faktorstufen steigen oder fallen. Bei zwei oder mehr Faktoren deuten parallele Linien an, daß es keine Wechselwirkung zwischen den Faktoren gibt. Das heißt, daß Sie die Faktorstufen eines einzelnen Faktors untersuchen können. Nichtparallele Linien deuten auf eine Wechselwirkung hin.

Abbildung 23-6
Nichtparalleles Diagramm (links) und paralleles Diagramm (rechts)



Nachdem ein Diagramm durch Auswahl von Faktoren für die horizontale Achse (und wahlweise von Faktoren für getrennte Linien und getrennte Diagramme) festgelegt wurde, muß das Diagramm der Liste "Diagramme" hinzugefügt werden.

GLM: Post-Hoc-Vergleiche

Abbildung 23-7
Dialogfeld "Univariat: Post-Hoc-Mehrfachvergleiche für beobachteten Mittelwert"

Tests für Post-Hoc-Mehrfachvergleiche. Sobald Sie festgestellt haben, daß es Abweichungen zwischen den Mittelwerten gibt, können Sie mit Post-Hoc-Spannweiten-Tests und paarweisen multiplen Vergleichen untersuchen, welche Mittelwerte sich unterscheiden. Die Vergleiche werden auf der Basis von nicht korrigierten Werten vorgenommen. Diese Tests werden nur für feste Zwischensubjektfaktoren durchgeführt. Bei "GLM - Meßwiederholungen" sind diese Tests nicht verfügbar, wenn es keine Zwischensubjektfaktoren gibt, und die Post-Hoc-Mehrfachvergleiche werden für den Durchschnitt aller Stufen der Innersubjektfaktoren durchgeführt. Bei "GLM - Multivariat" werden für jede abhängige Variable eigene Post-Hoc-Tests durchgeführt. "GLM – Multivariat" und "GLM – Meßwiederholungen" sind nur verfügbar, wenn Sie SPSS Advanced Models installiert haben.

Häufig verwendete Mehrfachvergleiche sind der Bonferroni-Test und die ehrlich signifikante Differenz nach Tukey. Der **Bonferroni-Test** auf der Grundlage der studentisierten *T*-Statistik korrigiert das beobachtete Signifikanzniveau unter Berücksichtigung der Tatsache, daß multiple Vergleiche vorgenommen werden. Der **Sidak-T-Test** korrigiert ebenfalls das Signifikanzniveau und liefert engere Grenzen als der Bonferroni-Test. Die **ehrlich signifikante Differenz nach Tukey** verwendet die studentisierte Spannweitenstatistik, um alle paarweisen Vergleiche zwischen den

Gruppen vorzunehmen, und setzt die experimentelle Fehlerrate auf die Fehlerrate der Ermittlung aller paarweisen Vergleiche. Beim Testen einer großen Anzahl von Mittelwertpaaren ist der Test auf ehrlich signifikante Differenz nach Tukey leistungsfähiger als der Bonferroni-Test. Bei einer kleinen Anzahl von Paaren ist der Bonferroni-Test leistungsfähiger.

GT2 nach Hochberg ähnelt dem Test auf ehrlich signifikante Differenz nach Tukey, es wird jedoch das studentisierte Maximalmodul verwendet. Meistens ist der Test nach Tukey leistungsfähiger. Der **paarweise Vergleichstest nach Gabriel** verwendet ebenfalls das studentisierte Maximalmodul und zeigt meistens eine größere Schärfe als das GT2 nach Hochberg, wenn die Zellengrößen ungleich sind. Der Gabriel-Test kann ungenau werden, wenn die Zellengrößen stark variieren.

Mit dem **paarweisen T-Test für mehrere Vergleiche nach Dunnett** wird ein Set von Verarbeitungen mit einem einzelnen Kontrollmittelwert verglichen. Als Kontrollkategorie ist die letzte Kategorie voreingestellt. Sie können aber auch die erste Kategorie einstellen. Außerdem können Sie einen einseitigen oder zweiseitigen Test wählen. Verwenden Sie einen zweiseitigen Test, um zu überprüfen, ob sich der Mittelwert bei jeder Stufe (außer der Kontrollkategorie) des Faktors von dem Mittelwert der Kontrollkategorie unterscheidet. Wählen Sie < Kontrolle, um zu überprüfen, ob der Mittelwert bei allen Stufen des Faktors kleiner als der Mittelwert der Kontrollkategorie ist. Wählen Sie > Kontrolle, um zu überprüfen, ob der Mittelwert bei allen Stufen des Faktors größer als der Mittelwert bei der Kontrollkategorie ist.

Ryan, Einot, Gabriel und Welsch (R-E-G-W) entwickelten zwei multiple Step-Down-Spannweitentests. Multiple Step-Down-Prozeduren überprüfen zuerst, ob alle Mittelwerte gleich sind. Wenn nicht alle Mittelwerte gleich sind, werden Teilmengen der Mittelwerte auf Gleichheit getestet. Das **F nach R-E-G-W** basiert auf einem *F*-Test, und **Q nach R-E-G-W** basiert auf der studentisierten Spannweite. Diese Tests sind leistungsfähiger als der multiple Spannweitentest nach Duncan und der Student-Newman-Keuls-Test (ebenfalls multiple Step-Down-Prozeduren), aber sie sind bei ungleichen Zellengrößen nicht empfehlenswert.

Bei ungleichen Varianzen verwenden Sie das **Tamhane-T2** (konservativer paarweiser Vergleichstest auf der Grundlage eines *T*-Tests), **Dunnett-T3** (paarweiser Vergleichstest auf der Grundlage des studentisierten Maximalmoduls), den **paarweisen Vergleichstest** nach **Games-Howell** (manchmal ungenau) oder das **Dunnett-C** (paarweiser Vergleichstest auf der Grundlage der studentisierten Spannweite).

Der **multiple Spannweitentest nach Duncan**, Student-Newman-Keuls (**S-N-K**) und **Tukey-B** sind Spannweitentests, mit denen Mittelwerte von Gruppen geordnet und ein Wertebereich berechnet wird. Diese Tests werden nicht so häufig verwendet wie die vorher beschriebenen Tests.

Der **Waller-Duncan-T-Test** verwendet die Bayes-Methode. Dieser Spannweitentest verwendet den harmonischen Mittelwert der Stichprobengröße, wenn die Stichprobengrößen ungleich sind.

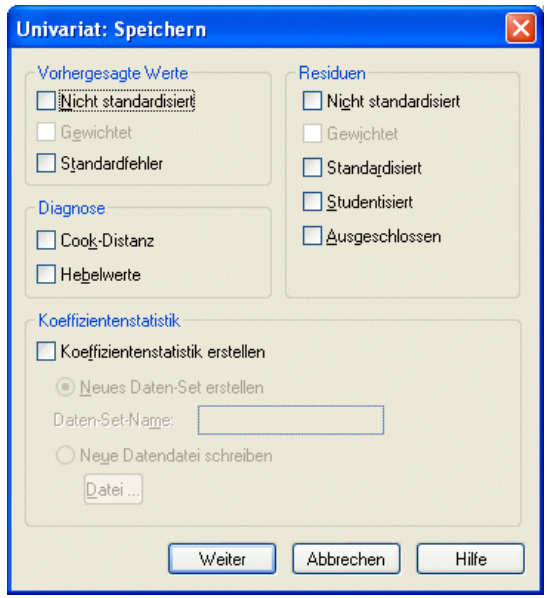
Das Signifikanzniveau des **Scheffé**-Tests ist so festgelegt, daß alle möglichen linearen Kombinationen von Gruppenmittelwerten getestet werden können und nicht nur paarweise Vergleiche verfügbar sind, wie bei dieser Funktion der Fall. Das führt dazu, daß der Scheffé-Test oftmals konservativer als andere Tests ist, also für eine Signifikanz eine größere Differenz der Mittelwerte erforderlich ist.

Der paarweise multiple Vergleichstest auf geringste signifikante Differenz (**LSD**) ist äquivalent zu multiplen individuellen *T*-Tests zwischen allen Gruppenpaaren. Der Nachteil bei diesem Test ist, daß kein Versuch unternommen wird, das beobachtete Signifikanzniveau im Hinblick auf multiple Vergleiche zu korrigieren.

Angezeigte Tests. Es werden paarweise Vergleiche für LSD, Sidak, Bonferroni, Games und Howell, T2 und T3 nach Tamhane, Dunnett-*C* und Dunnett-T3 ausgegeben. Homogene Untergruppen für Spannweitentests werden ausgegeben für S-N-K, Tukey-*B*, Duncan, *F* nach R-E-G-W, *Q* nach R-E-G-W und Waller. Die ehrlich signifikante Differenz nach Tukey, das GT2 nach Hochberg, der Gabriel-Test und der Scheffé-Test sind multiple Vergleiche, zugleich aber auch Spannweitentests.

GLM: Speichern

Abbildung 23-8
Dialogfeld "Univariat: Speichern"



Vom Modell vorhergesagte Werte, Residuen und verwandte Maße können als neue Variablen im Daten-Editor gespeichert werden. Viele dieser Variablen können zum Untersuchen von Annahmen über die Daten verwendet werden. Um die Werte zur Verwendung in einer anderen SPSS-Sitzung zu speichern, müssen Sie die aktuelle Datendatei speichern.

Vorhergesagte Werte. Dies sind die Werte, welche das Modell für jeden Fall vorhersagt.

- **Nicht standardisiert.** Der Wert, den das Modell für die abhängige Variable vorhersagt.
- **Gewichtet.** Gewichtete nichtstandardisierte vorhergesagte Werte. Nur verfügbar, wenn zuvor eine WLS-Variable ausgewählt wurde.
- **Standardfehler.** Ein Schätzwert der Standardabweichung des Durchschnittswertes der abhängigen Variablen für die Fälle, die denselben Werte für die unabhängigen Variablen haben.

Diagnose. Dies sind Maße zum Auffinden von Fällen mit ungewöhnlichen Wertekombinationen bei der unabhängigen Variablen und von Fällen, die einen großen Einfluß auf das Modell haben könnten.

- **Cook-Distanz.** Dies ist ein Maß dafür, wie stark sich die Residuen aller Fälle ändern würden, wenn ein spezieller Fall von der Berechnung der Regressionskoeffizienten ausgeschlossen würde. Ein großer Wert der Cook-Distanz zeigt an, daß der Ausschluß eines Falles von der Berechnung der Regressionskoeffizienten die Koeffizienten substantiell verändert.
- **Hebelwerte.** Nicht zentrierte Hebelwerte. Der relative Einfluß einer jeden Beobachtung auf die Anpassungsgüte eines Modells.

Residuen. Ein nicht standardisiertes Residuum ist der tatsächliche Wert der abhängigen Variablen minus dem vom Modell geschätzten Wert. Ebenfalls verfügbar sind standardisierte, studentisierte und ausgeschlossene Residuen. Falls Sie eine WLS-Variable ausgewählt haben, sind auch gewichtete nicht standardisierte Residuen verfügbar.

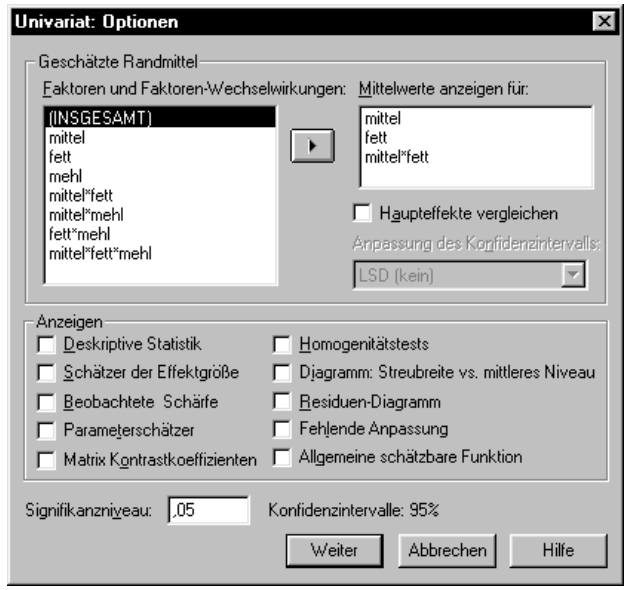
- **Nicht standardisiert.** Die Differenz zwischen einem beobachteten Wert und dem durch das Modell vorhergesagten Wert.
- **Gewichtet.** Gewichtete nichtstandardisierte Residuen. Nur verfügbar, wenn zuvor eine WLS-Variable ausgewählt wurde.
- **Standardisiert.** Der Quotient aus dem Residuum und einem Schätzer seines Standardfehlers. Standardisierte Residuen, auch bekannt als Pearson-Residuen, haben einen Mittelwert von 0 und eine Standardabweichung von 1.
- **Studentisiert.** Ein Residuum, das durch seine geschätzte Standardabweichung geteilt wird, die je nach der Distanz zwischen den Werten der unabhängigen Variablen des Falles und dem Mittelwert der unabhängigen Variablen von Fall zu Fall variiert.
- **Ausgeschlossen.** Das Residuum für einen Fall, wenn dieser Fall nicht in die Berechnung der Regressionskoeffizienten eingegangen ist. Es ist die Differenz zwischen dem Wert der abhängigen Variablen und dem korrigierten Schätzwert.

Koeffizientenstatistik. Hiermit wird eine Varianz-Kovarianz-Matrix der Parameterschätzungen für das Modell in ein neues Daten-Set in der aktuellen Sitzung oder in eine externe Datei im SPSS-Format geschrieben. Für jede abhängige Variable gibt es weiterhin eine Zeile mit Parameterschätzungen, eine Zeile mit Signifikanzwerten für die *T*-Statistik der betreffenden Parameterschätzungen und eine Zeile mit den Freiheitsgraden der Residuen. Bei multivariaten Modellen gibt es

ähnliche Zeilen für jede abhängige Variable. Sie können diese Matrixdatei auch in anderen Prozeduren verwenden, die SPSS-Matrixdateien einlesen.

GLM-Optionen

Abbildung 23-9
Dialogfeld "Univariat: Optionen"



In diesem Dialogfeld sind weitere Statistiken verfügbar. Diese werden auf der Grundlage eines Modells mit festen Effekten berechnet.

Geschätzte Randmittel. Wählen Sie die Faktoren und Wechselwirkungen aus, für die Sie Schätzer für die Randmittel der Grundgesamtheit in den Zellen wünschen. Diese Mittel werden gegebenenfalls an die Kovariaten angepaßt.

- **Haupteffekte vergleichen.** Gibt nicht korrigierte paarweise Vergleiche zwischen den geschätzten Randmitteln für alle Haupteffekte im Modell aus, sowohl für Zwischensubjektfaktoren als auch für Innersubjektfaktoren. Diese Option ist nur

verfügbar, falls in der Liste “Mittelwerte anzeigen für” Haupteffekte ausgewählt sind.

- **Anpassung des Konfidenzintervalls.** Wählen Sie für das Konfidenzintervall und die Signifikanz entweder die geringste signifikante Differenz (LSD; least significant difference), Bonferroni oder die Anpassung nach Sidak. Diese Option ist nur verfügbar, wenn Haupteffekte vergleichen ausgewählt ist.

Anzeigen. Mit der Option Deskriptive Statistik lassen Sie beobachtete Mittelwerte, Standardabweichungen und Häufigkeiten für alle abhängigen Variablen in allen Zellen berechnen. Die Option Schätzer der Effektgröße liefert einen partiellen Eta-Quadrat-Wert für jeden Effekt und jede Parameterschätzung. Die Eta-Quadrat-Statistik beschreibt den Anteil der Gesamtvariabilität, der einem Faktor zugeschrieben werden kann. Die Option Beobachtete Schärfe liefert die Testschärfe, wenn die alternative Hypothese auf die Basis der beobachteten Werte eingestellt wurde. Mit Parameterschätzer werden Parameterschätzer, Standardfehler, *T*-Tests, Konfidenzintervalle und die beobachtete Schärfe für jeden Test berechnet. Mit der Option Matrix-Kontrastkoeffizienten wird die **L**-Matrix berechnet.

Mit der Option Homogenitätstest wird der Levene-Test auf Homogenität der Varianzen für alle abhängigen Variablen über alle Kombinationen von Faktorstufen der Zwischensubjektfaktoren durchgeführt (nur für Zwischensubjektfaktoren). Die Optionen für Diagramme der Streubreite gegen das mittlere Niveau und Residuen-Diagramme sind beim Überprüfen von Annahmen über die Daten nützlich. Diese Option ist nur verfügbar, wenn Faktoren vorhanden sind. Wählen Sie Residuen-Diagramm, wenn Sie für jede abhängige Variable ein Residuen-Diagramm (beobachtete über vorhergesagte über standardisierte Werte) erhalten möchten. Diese Diagramme sind beim Überprüfen der Annahme von Gleichheit der Varianzen nützlich. Mit der Option Fehlende Anpassung können Sie überprüfen, ob das Modell die Beziehung zwischen der abhängigen Variablen und der unabhängigen Variablen richtig beschreiben kann. Die Option Allgemeine schätzbare Funktion ermöglicht Ihnen, einen benutzerdefinierten Hypothesentest zu entwickeln, dessen Grundlage die allgemeine schätzbare Funktion ist. Zeilen in einer beliebigen Matrix der Kontrastkoeffizienten sind lineare Kombinationen der allgemeinen schätzbaren Funktion.

Signifikanzniveau. Hier können Sie das in den Post-Hoc-Tests verwendete Signifikanzniveau und das beim Berechnen von Konfidenzintervallen verwendete Konfidenzniveau ändern. Der hier festgelegte Wert wird auch zum Berechnen der

beobachteten Schärfe für die Tests verwendet. Wenn Sie ein Signifikanzniveau festlegen, wird das entsprechende Konfidenzniveau im Dialogfeld angezeigt.

Zusätzliche Funktionen beim Befehl UNIANOVA

Mit der SPSS-Befehlssprache verfügen Sie über die folgenden zusätzlichen Möglichkeiten:

- Mit dem Unterbefehl `DESIGN` können Sie verschachtelte Effekte im Design festlegen.
- Mit dem Unterbefehl `TEST` können Sie Tests auf Effekte im Vergleich zu linearen Kombinationen von Effekten oder einem Wert vornehmen.
- Mit dem Unterbefehl `CONTRAST` können Sie multiple Kontraste angeben.
- Mit dem Unterbefehl `MISSING` können Sie benutzerdefinierte fehlende Werte aufnehmen.
- Mit dem Unterbefehl `CRITERIA` können Sie EPS-Kriterien angeben.
- Mit den Unterbefehlen `LMATRIX`, `MMATRIX` und `KMATRIX` können Sie benutzerdefinierte **L**-Matrizen, **M**-Matrizen und **K**-Matrizen erstellen.
- Mit dem Unterbefehl `CONTRAST` können Sie bei einfachen und Abweichungskontrasten eine Referenzkategorie zwischenschalten.
- Mit dem Unterbefehl `CONTRAST` können Sie bei polynomialen Kontrasten Metriken angeben.
- Mit dem Unterbefehl `POSTHOC` können Sie Fehlerterme für Post-Hoc-Vergleiche angeben.
- Mit dem Unterbefehl `EMMEANS` können Sie geschätzte Randmittel für alle Faktoren oder Faktorenwechselwirkungen zwischen den Faktoren in der Faktorenliste berechnen lassen.
- Mit dem Unterbefehl `SAVE` können Sie Namen für temporäre Variablen angeben.
- Mit dem Unterbefehl `OUTFILE` können Sie eine Datendatei mit einer Korrelationsmatrix erstellen.
- Mit dem Unterbefehl `OUTFILE` können Sie eine Matrix-Datendatei erstellen, die Statistiken aus der Zwischensubjekt-ANOVA-Tabelle enthält.
- Mit dem Unterbefehl `OUTFILE` können Sie die Design-Matrix in einer neuen Datendatei speichern.

Vollständige Informationen zur Syntax finden Sie im *SPSS Syntax Reference Guide*.

Kapitel
24

Bivariate Korrelationen

Mit der Prozedur "Bivariate Korrelationen" werden der Korrelationskoeffizient nach Pearson, Spearman-Rho und Kendall-Tau-*b* mit ihren jeweiligen Signifikanzniveaus errechnet. Mit Korrelationen werden die Beziehungen zwischen Variablen oder deren Rängen gemessen. Untersuchen Sie Ihre Daten vor dem Berechnen eines Korrelationskoeffizienten auf Ausreißer, da diese zu irreführenden Ergebnissen führen können. Stellen Sie fest, ob wirklich ein linearer Zusammenhang existiert. Der Korrelationskoeffizient nach Pearson ist ein Maß für den linearen Zusammenhang. Wenn zwei Variablen miteinander in starker Beziehung stehen, der Zusammenhang aber nicht linear ist, ist der Korrelationskoeffizient nach Pearson keine geeignete Statistik zum Messen des Zusammenhangs.

Beispiel. Besteht eine Korrelation zwischen der Anzahl der von einer Basketballmannschaft gewonnenen Spiele und der durchschnittlich pro Spiel erzielten Anzahl von Punkten? Ein Streudiagramm zeigt, daß ein linearer Zusammenhang besteht. Eine Analyse der Daten der NBA-Saison 1994–1995 ergibt, daß der Korrelationskoeffizient nach Pearson (0,581) auf dem Niveau 0,01 signifikant ist. Man könnte vermuten, daß die gegnerischen Mannschaften um so weniger Punkte erreicht haben, je mehr Spiele eine Mannschaft gewann. Zwischen diesen Variablen besteht eine negative Korrelation (–0,401), die auf dem Niveau 0,05 signifikant ist.

Statistiken. Für jede Variable: Anzahl der Fälle mit nichtfehlenden Werten, Mittelwert und Standardabweichung. Für jedes Variablenpaar: Korrelationskoeffizient nach Pearson, Spearman-Rho, Kendall-Tau-*b*, Kreuzprodukt der Abweichungen und Kovarianz.

Daten. Verwenden Sie symmetrische quantitative Variablen für den Korrelationskoeffizienten nach Pearson und quantitative Variablen oder Variablen mit ordinalskalierten Kategorien für das Spearman-Rho und Kendall-Tau-*b*.

Annahmen. Für den Korrelationskoeffizient nach Pearson wird angenommen, daß jedes Variablenpaar bivariat normalverteilt ist.

Abbildung 24-1
Ausgabe für die Prozedur "Bivariate Korrelationen"

Korrelationen

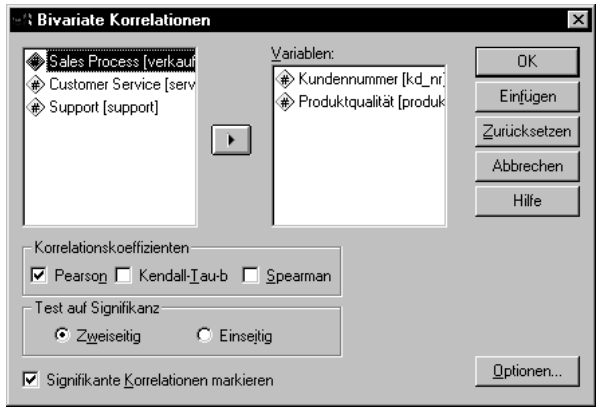
		Gewonnene Spiele	Eigene Punkte pro Spiel	Gegnerische Punkte pro Spiel
Korrelation nach Pearson	Gewonnene Spiele	1.000	.581	-.401
	Eigene Punkte pro Spiel	.581	1.000	.457
	Gegnerische Punkte pro Spiel	-.401	.457	1.000
Signifikanz (2-seitig)	Gewonnene Spiele	.	.001	.038
	Eigene Punkte pro Spiel	.001	.	.017
	Gegnerische Punkte pro Spiel	.038	.017	.
N	Gewonnene Spiele	27	27	27
	Eigene Punkte pro Spiel	27	27	27
	Gegnerische Punkte pro Spiel	27	27	27

So lassen Sie bivariate Korrelationen berechnen

Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Korrelation
    Bivariat…

Abbildung 24-2
Dialogfeld "Bivariate Korrelationen"



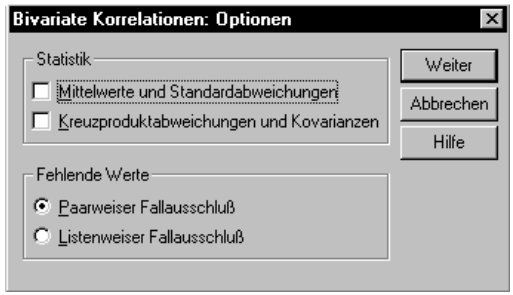
▶ Wählen Sie mindestens zwei numerische Variablen aus.

Außerdem sind folgende Optionen verfügbar:

- **Korrelationskoeffizienten.** Für quantitative, normalverteilte Variablen wählen Sie den Korrelationskoeffizienten nach Pearson. Wenn ihre Daten nicht normalverteilt sind oder mit geordneten Kategorien vorliegen, wählen Sie die Methoden Kendall-Tau-b oder Spearman, mit denen die Beziehungen zwischen Rangordnungen gemessen werden. Der Wertebereich für Korrelationskoeffizienten reicht von –1 (perfekter negativer Zusammenhang) bis +1 (perfekter positiver Zusammenhang). Der Wert 0 bedeutet, daß kein linearer Zusammenhang besteht. Vermeiden Sie bei der Interpretation Ihrer Ergebnisse, Schlüsse über Ursache und Wirkung aufgrund signifikanter Korrelationen zu ziehen.
- **Test auf Signifikanz.** Sie können einseitige oder zweiseitige Wahrscheinlichkeiten wählen. Wenn Ihnen die Richtung des Zusammenhangs im voraus bekannt ist, wählen Sie Einseitig. Wählen Sie anderenfalls Zweiseitig.
- **Signifikante Korrelationen markieren.** Korrelationskoeffizienten, die signifikant auf dem 0,05-Niveau liegen, werden mit einem einfachen Stern angezeigt. Liegen diese signifikant auf dem 0,01-Niveau, werden sie mit zwei Sternen angezeigt.

Bivariate Korrelationen: Optionen

Abbildung 24-3
Dialogfeld "Bivariate Korrelationen: Optionen"



Statistik. Für Pearson-Korrelationen können Sie eine oder auch beide der folgenden Optionen wählen:

- **Mittelwerte und Standardabweichungen.** Diese werden für jede Variable angezeigt. Außerdem wird die Anzahl der Fälle mit nichtfehlenden Werten angezeigt. Fehlende Werte werden Variable für Variable bearbeitet, unabhängig von Ihren Einstellungen für fehlende Werte.
- **Kreuzproduktabweichungen und Kovarianzen.** Werden für jedes Variablenpaar angezeigt. Das Kreuzprodukt der Abweichungen ist gleich der Summe der Produkte mittelwertkorrigierter Variablen. Dies ist der Zähler des Korrelationskoeffizienten nach Pearson. Die Kovarianz ist ein nicht standardisiertes Maß für den Zusammenhang zwischen zwei Variablen und ist gleich der Kreuzproduktabweichung dividiert durch *N*–1.

Fehlende Werte. Sie können eine der folgenden Optionen auswählen:

- **Paarweiser Fallausschluß.** Fälle mit fehlenden Werten für eine oder beide Variablen eines Paares für einen Korrelationskoeffizienten werden von der Analyse ausgeschlossen. Da jeder Koeffizient auf allen Fällen mit gültigen Codes für dieses bestimmte Variablenpaar basiert, werden in allen Berechnungen die maximal zugänglichen Informationen verwendet. Dies kann zu einer Menge von Koeffizienten führen, die auf einer variierenden Anzahl von Fällen basiert.
- **Listenweiser Fallausschluß.** Fälle mit fehlenden Werten für Variablen werden von allen Korrelationen ausgeschlossen.

Zusätzliche Funktionen bei den Befehlen CORRELATIONS und NONPAR CORR

Mit der SPSS-Befehlssprache verfügen Sie über die folgenden zusätzlichen Möglichkeiten:

- Mit dem Unterbefehl `MATRIX` kann eine Korrelationsmatrix für Pearson-Korrelationen geschrieben werden. Diese kann anstelle von Rohdaten verwendet werden, um andere Analysen zu berechnen, beispielsweise die Faktorenanalyse.
- Mit dem Schlüsselwort `WITH` im Unterbefehl `VARIABLES` können die Korrelationen zwischen allen Variablen einer Liste und allen Variablen einer zweiten Liste berechnet werden.

Vollständige Informationen zur Syntax finden Sie im *SPSS Syntax Reference Guide*.

Kapitel
25

Partielle Korrelationen

Partielle Korrelationskoeffizienten beschreiben die Beziehung zwischen zwei Variablen. Die Prozedur "Partielle Korrelationen" berechnet diese Koeffizienten, wobei die Effekte von einer oder mehr zusätzlichen Variablen überprüft werden. Korrelationen sind Maße für lineare Zusammenhänge. Zwei Variablen können fehlerlos miteinander verbunden sein. Wenn es sich aber nicht um eine lineare Beziehung handelt, ist der Korrelationskoeffizient zur Messung des Zusammenhangs zwischen den beiden Variablen nicht geeignet.

Beispiel. Besteht eine Beziehung zwischen den Ausgaben für das Gesundheitswesen und den Krankheitsraten? Obwohl man annehmen könnte, eine solche Beziehung sei negativ, ergibt eine Studie eine signifikante *positive* Korrelation: mit ansteigenden Ausgaben im Gesundheitswesen scheinen die Krankheitsraten zuzunehmen. Durch die Kontrolle der Effekte aus der Häufigkeit der Besuche bei medizinischem Personal wird die beobachtete positive Korrelation praktisch eliminiert. Die Ausgaben im Gesundheitswesen und die Krankheitsraten scheinen lediglich in einer positiven Beziehung zu stehen, da mit steigender Finanzausstattung mehr Menschen Zugang zu medizinischer Versorgung haben, was zu mehr gemeldeten Krankheiten bei Ärzten und Krankenhäusern führt.

Statistiken. Für jede Variable: Anzahl der Fälle mit nichtfehlenden Werten, Mittelwert und Standardabweichung. Matrizen für partielle Korrelationen und Korrelationen nullter Ordnung mit Freiheitsgraden und Signifikanzniveaus.

Daten. Verwenden Sie symmetrische, quantitative Variablen.

Annahmen. Die Prozedur "Partielle Korrelation" setzt für jedes Variablenpaar eine bivariate Normalverteilung voraus.

Abbildung 25-1
Ausgabe der partiellen Korrelationen

Correlations

Control Variables			Health care funding (amount per 100)	Reported diseases (rate per 10,000)	Visits to health care providers (rate per 10,000)
-none- [1]	Health care funding (amount per 100)	Correlation	1.000	.737	.964
		Significance (2-tailed)	.	.000	.000
		df	0	48	48
	Reported diseases (rate per 10,000)	Correlation	.737	1.000	.762
		Significance (2-tailed)	.000	.	.000
		df	48	0	48
	Visits to health care providers (rate per 10,000)	Correlation	.964	.762	1.000
		Significance (2-tailed)	.000	.000	.
		df	48	48	0
Visits to health care providers (rate per 10,000)	Health care funding (amount per 100)	Correlation	1.000	.013	
		Significance (2-tailed)	.	.928	
		df	0	47	
	Reported diseases (rate per 10,000)	Correlation	.013	1.000	
		Significance (2-tailed)	.928	.	
		df	47	0	

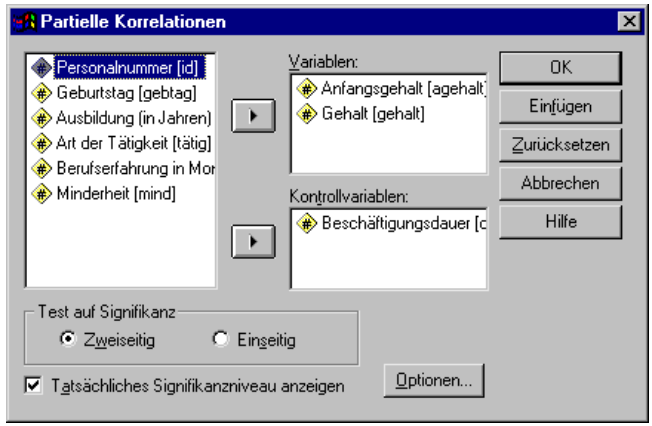
1. Cells contain zero-order (Pearson) correlations.

So lassen Sie partielle Korrelationen berechnen

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
Korrelation
Partiell...

Abbildung 25-2
Dialogfeld "Partielle Korrelationen"



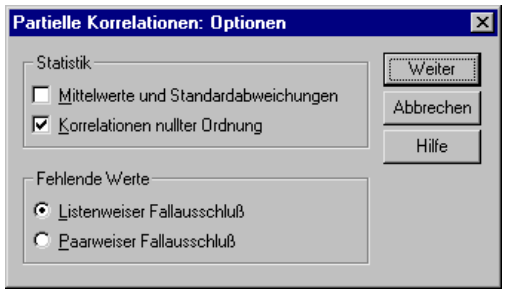
- Wählen Sie mindestens zwei numerische Variablen aus, für die partielle Korrelationen berechnet werden sollen.
- Wählen Sie mindestens eine numerische Kontroll-Variable aus.

Außerdem sind folgende Optionen verfügbar:

- **Test auf Signifikanz.** Sie können einseitige oder zweiseitige Wahrscheinlichkeiten wählen. Wenn Ihnen die Richtung des Zusammenhangs im voraus bekannt ist, wählen Sie Einseitig. Wählen Sie anderenfalls Zweiseitig.
- **Tatsächliches Signifikanzniveau anzeigen.** In der Standardeinstellung werden die Wahrscheinlichkeit sowie die Freiheitsgrade für jeden Korrelationskoeffizienten angezeigt. Wenn Sie diese Option deaktivieren, werden die Koeffizienten mit einem Signifikanzniveau von 0,05 mit einem Sternchen gekennzeichnet. Koeffizienten mit einem Signifikanzniveau von 0,01 werden mit einem doppelten Sternchen gekennzeichnet, und Freiheitsgrade werden unterdrückt. Diese Einstellung beeinflußt sowohl die Matrizen der partiellen Korrelationen als auch die der nullten Ordnung.

Partielle Korrelationen: Optionen

Abbildung 25-3
Dialogfeld "Partielle Korrelationen: Optionen"



Statistik. Sie können eine oder beide der folgenden Möglichkeiten auswählen:

- **Mittelwerte und Standardabweichungen.** Diese werden für jede Variable angezeigt. Außerdem wird die Anzahl der Fälle mit nichtfehlenden Werten angezeigt.
- **Korrelationen nullter Ordnung.** Hiermit wird eine einfache Matrix für Korrelationen zwischen allen Variablen (einschließlich Kontroll-Variablen) angezeigt.

Fehlende Werte. Sie können eine der folgenden Möglichkeiten wählen:

- **Listenweiser Fallausschluß.** Fälle mit fehlenden Werten für Variablen (einschließlich Kontroll-Variablen) werden aus den Berechnungen ausgeschlossen.
- **Paarweiser Fallausschluß.** Bei der Berechnung der Korrelationen nullter Ordnung, die den partiellen Korrelationen zugrunde liegen, werden Fälle mit fehlenden Werten in einer oder beiden Variablen eines Variablenpaars nicht verwendet. Beim paarweisen Löschen wird der größtmögliche Teil der Daten verwendet. Die Anzahl der Fälle kann jedoch von Koeffizient zu Koeffizient variieren. Wenn das paarweise Löschen aktiviert ist, liegt den Freiheitsgraden eines bestimmten partiellen Koeffizienten die niedrigste Anzahl von Fällen zugrunde, die zur Berechnung einer der Korrelationen nullter Ordnung verwendet werden.

Zusätzliche Funktionen beim Befehl PARTIAL CORR

Mit der SPSS-Befehlssprache verfügen Sie über die folgenden zusätzlichen Möglichkeiten:

- Sie können eine Korrelationsmatrix nullter Ordnung einlesen und eine Matrix der partiellen Korrelationen schreiben (mit dem Unterbefehl `MATRIX`).
- Sie können partielle Korrelationen zwischen zwei Variablenlisten erstellen (mit dem Schlüsselwort `WITH` im Unterbefehl `VARIABLES`).
- Sie können mehrere Analysen berechnen lassen (mit mehren Unterbefehlen `VARIABLES`).
- Sie können die Ordnung für die Anfrage angeben (z. B. partielle Korrelationen sowohl erster als auch zweiter Ordnung), wenn Sie über zwei Kontrollvariablen verfügen (mit dem Unterbefehl `VARIABLES`).
- Sie können redundante Koeffizienten unterdrücken (mit dem Unterbefehl `FORMAT`).
- Sie können eine Matrix von einfachen Korrelationen anzeigen lassen, wenn einige Koeffizienten nicht berechnet werden können (mit dem Unterbefehl `STATISTICS`).

Vollständige Informationen zur Syntax finden Sie im *SPSS Syntax Reference Guide*.

Kapitel
26

Distanzen

Durch diese Prozedur kann eine Vielzahl von Statistiken berechnet werden, indem Ähnlichkeiten oder Unähnlichkeiten (Distanzen) zwischen Paaren von Variablen oder Fällen gemessen werden. Diese Ähnlichkeits- oder Distanzmaße können dann bei anderen Prozeduren, beispielsweise der Faktorenanalyse, der Cluster-Analyse oder der multidimensionalen Skalierung zur Analyse komplexer Datenmengen verwendet werden.

Beispiel. Ist es möglich, Ähnlichkeiten zwischen Paaren von Kraftfahrzeugen anhand bestimmter Merkmale zu messen, z. B. anhand des Hubraums, des Kraftstoffverbrauchs oder der Leistung? Durch die Berechnung von Ähnlichkeiten zwischen Kraftfahrzeugen können Sie besser einordnen, welche Fahrzeuge einander ähneln bzw. welche sich voneinander unterscheiden. Mit einer hierarchischen Cluster-Analyse oder einer multidimensionalen Skalierung auf die Ähnlichkeiten können Sie eine formale Analyse durchführen, um die zugrundeliegende Struktur zu untersuchen.

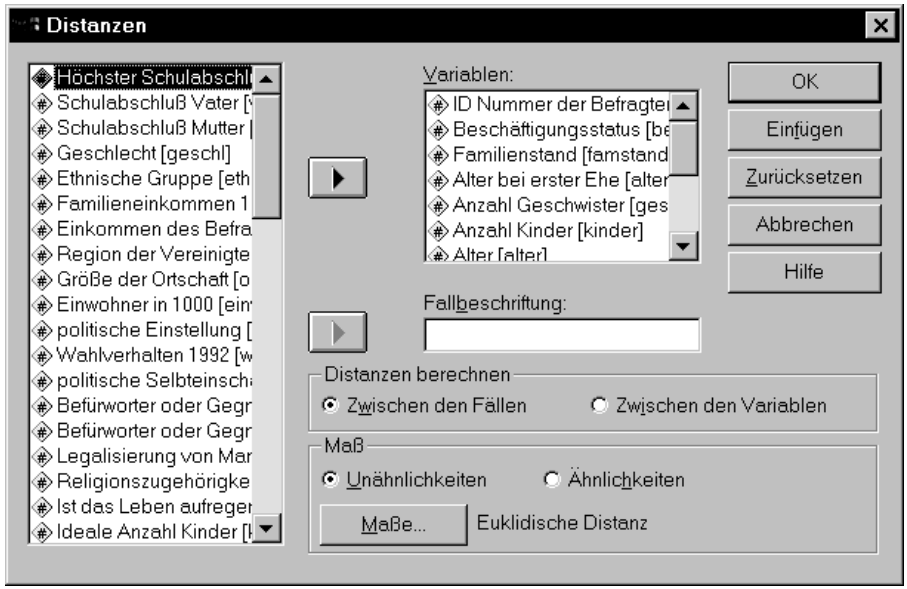
Statistiken. Unähnlichkeitsmaße (Distanzmaße) für Intervalldaten: Euklidischer Abstand, quadrierter Euklidischer Abstand, Tschebyscheff, Block, Minkowski oder ein benutzerdefiniertes Maß; für Häufigkeiten: Chi-Quadrat-Maß oder Phi-Quadrat-Maß; für Binärdaten: Euklidischer Abstand, quadrierter Euklidischer Abstand, Größendifferenz, Musterdifferenz, Varianz, Form und Distanzmaß nach Lance und Williams. Ähnlichkeitsmaße für Intervalldaten: Pearson-Korrelation oder Kosinus; für Binärdaten: Russel und Rao, einfache Übereinstimmung, Jaccard, Würfel-Ähnlichkeitsmaß, Ähnlichkeitsmaß nach Rogers und Tanimoto, Sokal und Sneath 1, Sokal und Sneath 2, Sokal und Sneath 3, Kulczynski 1, Kulczynski 2, Sokal und Sneath 4, Hamann, Lambda, Anderberg-*D*, Yule-*Y*, Yule-*Q*, Ochiai, Sokal und Sneath 5, Phi-4-Punkt-Korrelation oder Streuung.

So lassen Sie Distanzmatrizen berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Korrelation
    Distanzen...

Abbildung 26-1
Dialogfeld "Distanzen"

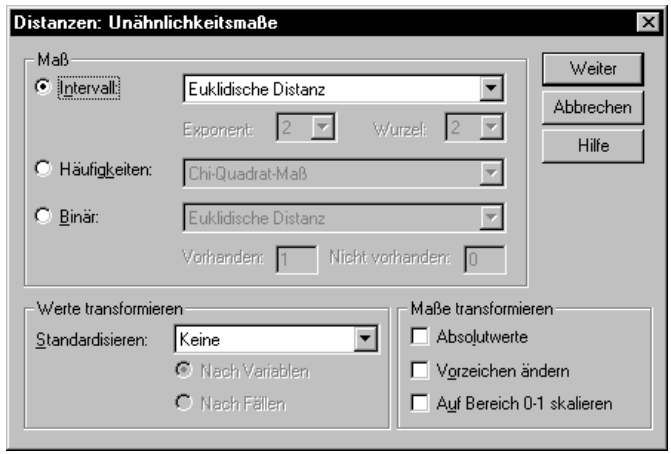


▶ Wählen Sie mindestens eine numerische Variable zur Berechnung von Distanzen zwischen Fällen, oder wählen Sie mindestens zwei numerische Variablen zur Berechnung von Distanzen zwischen Variablen.

▶ Wählen Sie im Gruppenfeld "Distanzen berechnen" eine andere Option aus, um Ähnlichkeiten zwischen Fällen oder Variablen zu berechnen.

Unähnlichkeitsmaße für Distanzen

Abbildung 26-2
Dialogfeld "Distanzen: Unähnlichkeitsmaße"



Wählen Sie aus dem Gruppenfeld "Maß" die Option aus, die Ihrem Datentyp entspricht ("Intervall", "Häufigkeiten" oder "Binär"). Wählen Sie dann aus dem Dropdown-Listenfeld ein Maß aus, das diesem Datentyp entspricht. Die folgenden Maße sind je nach Datentyp verfügbar:

- **Intervall.** Euklidischer Abstand, quadrierter Euklidischer Abstand, Tschebyscheff, Block, Minkowski oder ein benutzerdefiniertes Maß.
- **Häufigkeiten.** Chi-Quadrat-Maß oder Phi-Quadrat-Maß.
- **Binär.** Euklidischer Abstand, quadrierter Euklidischer Abstand, Größendifferenz, Musterdifferenz, Varianz, Form und Distanzmaß nach Lance und Williams. (Geben Sie Werte in die Felder "Vorhanden" und "Nicht vorhanden" ein, um anzugeben, welche beiden Werte sinnvoll sind; alle übrigen Werte werden durch die Distanzmaße ignoriert.)

Im Gruppenfeld "Werte transformieren" können Sie festlegen, ob die Datenwerte für Fälle oder Werte *vor* dem Berechnen von Ähnlichkeiten für Fälle oder Variablen standardisiert werden. Diese Transformationen sind nicht auf binäre Daten anwendbar. Die verfügbaren Standardisierungsmethoden sind *Z*-Werte, Bereich –1 bis 1, Bereich 0 bis 1, maximale Größe von 1, Mittelwert gleich 1 und Standardabweichung gleich 1.

Im Gruppenfeld “Maße transformieren” können Sie festlegen, ob die durch das Distanzmaß erzeugten Werte transformiert werden. Dies erfolgt, nachdem das Distanzmaß berechnet wurde. Zu den verfügbaren Optionen zählen Absolutwerte, Ändern des Vorzeichens und Skalieren auf den Bereich 0–1.

Ähnlichkeitsmaße für Distanzen

Abbildung 26-3
Dialogfeld “Distanzen: Ähnlichkeitsmaße”

Wählen Sie aus dem Gruppenfeld “Maß” die Option aus, die Ihrem Datentyp entspricht (“Intervall” oder “Binär”). Wählen Sie dann aus dem Dropdown-Listenfeld ein Maß aus, das diesem Datentyp entspricht. Die folgenden Maße sind je nach Datentyp verfügbar:

- **Intervall.** Pearson-Korrelation oder Kosinus
- **Binär.** Russel und Rao, einfache Übereinstimmung, Jaccard, Würfel-Ähnlichkeitsmaß, Ähnlichkeitsmaß nach Rogers und Tanimoto, Ähnlichkeitsmaße nach Sokal und Sneath 1 bis 5, Kulczynski 1, Kulczynski 2, Sokal und Sneath 4, Hamann, Lambda, Anderberg-*D*, Yule-*Y*, Yule-*Q*, Ochiai, Sokal und Sneath 5, Phi-4-Punkt-Korrelation oder Streuung. (Geben Sie Werte in die Felder “Vorhanden” und “Nicht vorhanden” ein, um anzugeben, welche beiden Werte sinnvoll sind; alle übrigen Werte werden durch die Distanzmaße ignoriert.)

Im Gruppenfeld “Werte transformieren” können Sie festlegen, ob die Datenwerte für Fälle oder Variablen vor dem Berechnen von Ähnlichkeiten standardisiert werden. Diese Transformationen sind nicht auf binäre Daten anwendbar. Die verfügbaren

Standardisierungsmethoden sind *Z*-Werte, Bereich –1 bis 1, Bereich 0 bis 1, maximale Größe von 1, Mittelwert gleich 1 und Standardabweichung gleich 1.

Im Gruppenfeld "Maße transformieren" können Sie festlegen, ob die durch das Distanzmaß erzeugten Werte transformiert werden. Dies erfolgt, nachdem das Distanzmaß berechnet wurde. Zu den verfügbaren Optionen zählen Absolutwerte, Ändern des Vorzeichens und Skalieren auf den Bereich 0–1.

Zusätzliche Funktionen beim Befehl PROXIMITIES

In der Prozedur "Distanzen" wird die Befehlssyntax von `PROXIMITIES` verwendet. Mit der SPSS-Befehlssprache verfügen Sie über die folgenden zusätzlichen Möglichkeiten:

- Angeben einer Ganzzahl als Exponent für das Minkowski-Distanzmaß.
- Angeben von beliebigen Ganzzahlen als Exponent und Wurzel für ein benutzerdefiniertes Distanzmaß.

Vollständige Informationen zur Syntax finden Sie im *SPSS Syntax Reference Guide*.

Kapitel
27

Lineare Regression

Mit “Lineare Regression” werden die Koeffizienten der linearen Gleichung unter Einbeziehung einer oder mehrerer unabhängiger Variablen geschätzt, die den Wert der abhängigen Variablen am besten vorhersagen. Sie können beispielsweise den Versuch unternehmen, die Jahresverkaufsbilanz eines Verkäufers (die abhängige Variable) nach unabhängigen Variablen wie Alter, Bildungsstand und Anzahl der Berufsjahre vorherzusagen.

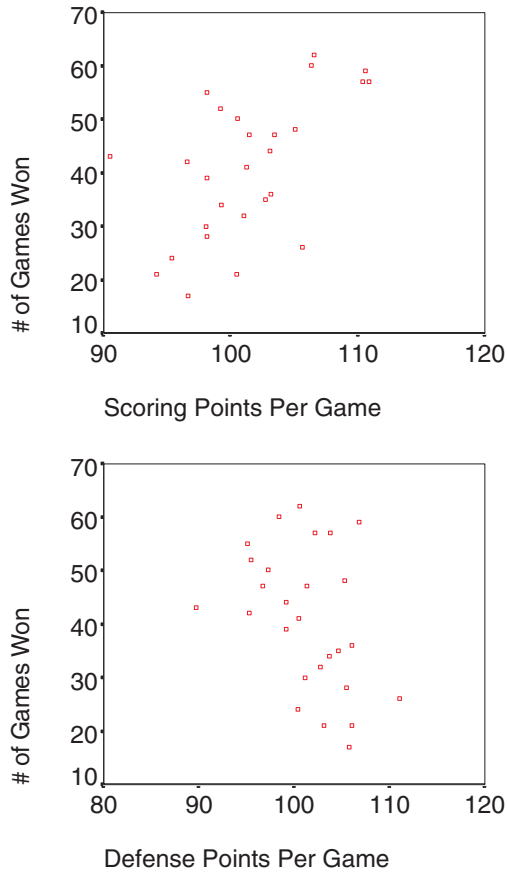
Beispiel. Besteht ein Zusammenhang zwischen der Anzahl der in einer Saison gewonnenen Spiele eines Basketball-Teams und der pro Spiel erzielten mittleren Punktezahl des Teams? Einem Streudiagramm läßt sich entnehmen, daß zwischen diesen Variablen eine lineare Beziehung besteht. Die Anzahl gewonnener Spiele und die erzielte Punktezahl des Gegners stehen gleichfalls in linearer Beziehung zueinander. Diese Variablen enthalten eine negative Beziehung. Einer steigenden Anzahl gewonnener Spiele steht eine fallende mittlere Punktezahl des Gegners gegenüber. Mit der linearen Regression können Sie die Beziehung dieser Variablen modellieren. Mit einem geeigneten Modell lassen sich Spielgewinne von Teams vorhersagen.

Statistiken. Für jede Variable: Anzahl gültiger Fälle, Mittelwert und Standardabweichung. Für jedes Modell: Regressionskoeffizienten, Korrelationsmatrix, Teil- und partielle Korrelationen, multiples R, R^2, korrigiertes R^2, Änderung in R^2, Standardfehler der Schätzung, Tabelle der Varianzanalyse, vorhergesagte Werte und Residuen. Außerdem 95%-Konfidenzintervalle für jeden Regressionskoeffizienten, Varianz-Kovarianz-Matrix, Inflationsfaktor der Varianz, Toleranz, Durbin-Watson-Test, Distanzmaße (Mahalanobis, Cook und Hebelwerte), DfBeta, DfFit, Vorhersageintervalle und fallweise Diagnose. Diagramme: Streudiagramme, partielle Diagramme, Histogramme und Normalverteilungsdiagramme.

Daten. Die abhängigen und die unabhängigen Variablen müssen quantitativ sein. Kategoriale Variablen, wie beispielsweise Religion, Studienrichtung oder Wohnsitz, müssen in binäre (Dummy-)Variablen oder andere Typen von Kontrast-Variablen umkodiert werden.

Annahmen. Für jeden Wert der unabhängigen Variablen muß die abhängige Variable normalverteilt vorliegen. Die Varianz der Verteilung der abhängigen Variablen muß für alle Werte der unabhängigen Variablen konstant sein. Die Beziehung zwischen der abhängigen Variablen und allen unabhängigen Variablen sollte linear sein, und alle Beobachtungen sollten unabhängig sein.

Abbildung 27-1
Ausgabe der linearen Regression



Modellzusammenfassung[2]

		R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
Modell	1	.947[1]	.898	.889	4.40

1. Einflußvariablen : (Konstante), Gegnerische Punkte pro Spiel, Eigene Punkte pro Spiel
2. Abhängige Variable

Koeffizienten[1]

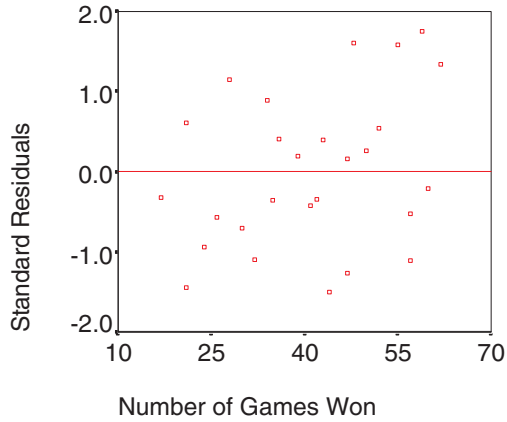
			Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten		
			B	Standardfehler	Beta	T	Signifikanz
Modell	1	(Konstante)	28.121	21.404		1.314	.201
		Eigene Punkte pro Spiel	2.539	.193	.965	13.145	.000
		Gegnerische Punkte pro Spiel	-2.412	.211	-.841	-11.458	.000

1. Abhängige Variable

Koeffizienten[1]

		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten		
Modell		B	Standardfehler	Beta	t	Sig.
1	(Konstante)	28.121	21.404		1.314	.201
	Erreichte Punkte pro Spiel	2.539	.193	.965	13.145	.000
	Abwehrpunkte pro Spiel	-2.412	.211	-.841	-11.458	.000

1. Abhängige Variable: Anzahl der gewonnenen Spiele

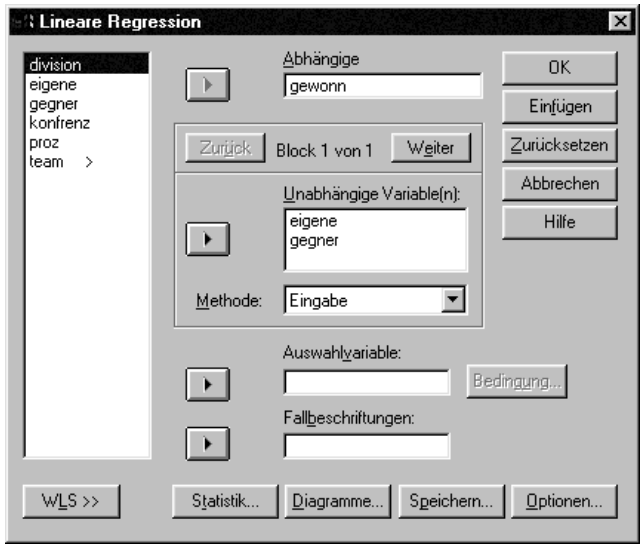


So lassen Sie eine lineare Regressionsanalyse berechnen

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Regression
    Linear...

Abbildung 27-2
Dialogfeld "Lineare Regression"



- Wählen Sie im Dialogfeld “Lineare Regression” eine numerische abhängige Variable aus.
- Wählen Sie eine oder mehrere numerische unabhängige Variablen aus.

Die folgenden Optionen sind verfügbar:

- Unabhängige Variablen können in Blöcken zusammengefaßt werden, und es können verschiedene Einschlußmethoden für unterschiedliche Untergruppen von Variablen angegeben werden.
- Auswahlvariablen zum Begrenzen der Analyse auf eine Untergruppe von Fällen mit einem bestimmten Wert oder bestimmten Werten für diese Variable können ausgewählt werden.
- Es können Variablen zur Fallunterscheidung ausgewählt werden, um Punkte in Diagrammen zu identifizieren.
- Wählen Sie eine numerische Variable für die WLS-Gewichtung aus, um eine Analyse der gewichteten kleinsten Quadrate durchzuführen.

Methode der gewichteten kleinsten Quadrate (WLS). Hiermit können Sie ein Modell gewichteter kleinster Quadrate berechnen. Die Datenpunkte werden mit dem reziproken Wert ihrer Varianzen gewichtet. Dies bedeutet, daß Beobachtungen mit großen Varianzen die Analyse weniger beeinflussen als Beobachtungen mit kleinen Varianzen.

Lineare Regression: Methode zur Auswahl von Variablen

Durch die Auswahl der Methode können Sie festlegen, wie unabhängige Variablen in die Analyse eingeschlossen werden. Anhand verschiedener Methoden können Sie eine Vielfalt von Regressionsmodellen mit demselben Satz von Variablen erstellen.

- **Einschluß (Regression).** Eine Prozedur für die Variablenauswahl, bei der alle Variablen eines Blocks in einem einzigen Schritt aufgenommen werden.
- **Schrittweise Selektion.** Bei jedem Schritt wird die noch nicht in der Gleichung enthaltene unabhängige Variable mit der kleinsten F-Wahrscheinlichkeit aufgenommen, sofern diese Wahrscheinlichkeit klein genug ist. Bereits in der Regressionsgleichung enthaltene Variablen werden entfernt, wenn ihre F-Wahrscheinlichkeit hinreichend groß wird. Das Verfahren endet, wenn keine Variablen mehr für Aufnahme oder Ausschluss in Frage kommen.

- **Entfernen (Variable Selection).** Ein Verfahren zur Variablenselektion, bei dem alle Variablen eines Blocks in einem Schritt ausgeschlossen werden.
- **Rückwärtselimination.** Eine Methode zur Variablendselektion, bei der alle Variablen in die Gleichung aufgenommen und anschließend sequentiell ausgeschlossen werden. Die Variable mit der kleinsten Teilkorrelation zur abhängigen Variablen wird als erste für den Ausschluß in Betracht gezogen. Wenn Sie das Ausschlußkriterium erfüllt, wird sie entfernt. Nach dem Ausschluß der ersten Variablen wird die nächste Variable mit der kleinsten Teilkorrelation in Betracht gezogen. Das Verfahren wird beendet, wenn keine Variablen mehr zur Verfügung stehen, die die Ausschlußkriterien erfüllen.
- **Vorwärtsselektion.** Ein Verfahren zur schrittweisen Variablenauswahl, in dem die Variablen nacheinander in das Modell aufgenommen werden. Die erste Variable, die in Betracht gezogen wird, ist die mit der größten positiven bzw. negativen Korrelation mit der abhängigen Variablen. Diese Variable wird nur dann in die Gleichung aufgenommen, wenn sie das Aufnahmekriterium erfüllt. Wenn die erste Variable ausgewählt ist, wird die unabhängige Variable mit der größten partiellen Korrelation betrachtet. Das Verfahren endet, wenn keine verbliebene Variable das Aufnahmekriterium erfüllt.

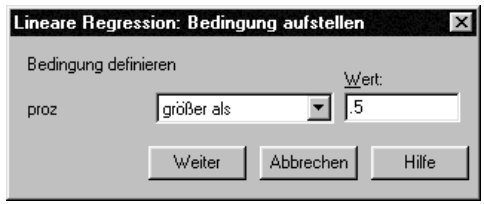
Die Signifikanzwerte in Ihrer Ausgabe basieren auf der Berechnung eines einzigen Modells. Deshalb sind diese generell ungültig, wenn eine schrittweise Methode (Schrittweise, Vorwärts oder Rückwärts) verwendet wird.

Alle Variablen müssen das Toleranzkriterium erfüllen, um unabhängig von der angegebenen Einschlußmethode in die Gleichung einbezogen zu werden. In der Standardeinstellung beträgt der Toleranzwert 0,0001. Eine Variable wird auch dann nicht eingeschlossen, wenn dadurch die Toleranz einer Variablen im Modell unter das Toleranzkriterium abfallen würde.

Alle ausgewählten unabhängigen Variablen werden einem einzigen Regressionsmodell hinzugefügt. Sie können jedoch verschiedene Einschlußmethoden für unterschiedliche Untergruppen von Variablen angeben. Beispielsweise können Sie einen Block von Variablen durch schrittweises Auswählen und einen zweiten Block durch Vorwärtsselektion in das Regressionsmodell einschließen. Um einem Regressionsmodell einen zweiten Block von Variablen hinzuzufügen, klicken Sie auf Weiter.

Lineare Regression: Bedingung aufstellen

Abbildung 27-3
Dialogfeld "Lineare Regression: Bedingung aufstellen"



Die durch die Auswahlbedingung definierten Fälle werden in die Analyse eingeschlossen. Wenn Sie für die Variable beispielsweise gleich wählen und als Wert 5 eingeben, werden nur Fälle in die Analyse einbezogen, für die der Wert der gewählten Variablen gleich 5 ist. Ein String-Wert ist ebenfalls möglich.

Lineare Regression: Diagramme

Abbildung 27-4
Dialogfeld "Lineare Regression: Diagramme"

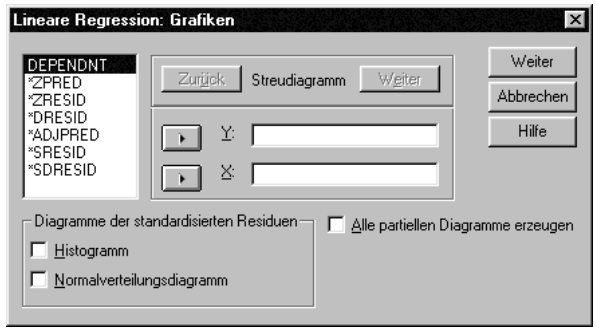


Diagramme können beim Validieren der Annahmen von Normalverteilung, Linearität und Varianz-Gleichheit hilfreich sein. Diagramme dienen auch zum Auffinden von Ausreißern, ungewöhnlichen Beobachtungen und Einflußfällen. Nachdem sie als neue Variablen gespeichert wurden, stehen im Daten-Editor vorhergesagte Werte,

Residuen und andere diagnostische Hilfsmittel zum Erstellen von Diagrammen mit den unabhängigen Variablen zur Verfügung. Folgende Diagramme sind verfügbar:

Streudiagramme. Sie können je zwei der folgenden Elemente auftragen: die abhängige Variable, standardisierte vorhergesagte Werte, standardisierte Residuen, ausgeschlossene Residuen, korrigierte vorhergesagte Werte, studentisierte Residuen oder studentisierte ausgeschlossene Residuen. Tragen Sie die standardisierten Residuen über den standardisierten vorhergesagten Werten auf, um auf Linearität und Varianz-Gleichheit zu überprüfen.

Liste der Quellvariablen. Listet die abhängigen Variablen (DEPENDNT) und die folgenden vorhergesagten und Residuen-Variablen auf: standardisierte vorhergesagte Werte (*ZPRED), standardisierte Residuen (*ZRESID), ausgeschlossene Residuen (*DRESID), korrigierte vorhergesagte Werte (*ADJPRED), studentisierte Residuen (*SRESID) und studentisierte ausgeschlossene Residuen (*SDRESID).

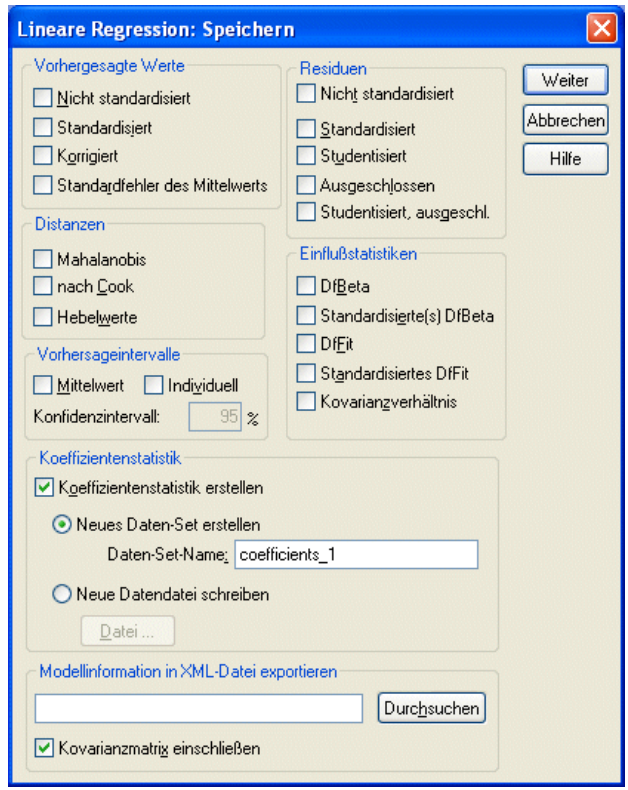
Alle partiellen Diagramme erzeugen. Erzeugt Streudiagramme der Residuen aller unabhängigen Variablen und der Residuen der abhängigen Variablen, wenn für den Rest der unabhängigen Variablen beide Variablen einer getrennten Regression unterzogen werden. Zum Erzeugen eines partiellen Diagramms müssen mindestens zwei unabhängige Variablen in der Gleichung enthalten sein.

Diagramme der standardisierten Residuen. Sie können Histogramme standardisierter Residuen und Normalverteilungsdiagramme anfordern, welche die Verteilung standardisierter Residuen mit einer Normalverteilung vergleichen.

Beim Anfordern von Diagrammen werden Auswertungsstatistiken für standardisierte vorhergesagte Werte und standardisierte Residuen (*ZPRED und *ZRESID) angezeigt.

Lineare Regression: Speichern von neuen Variablen

Abbildung 27-5
Dialogfeld "Lineare Regression: Speichern"



Vorhergesagte Werte, Residuen und andere für die Diagnose nützliche Statistiken können gespeichert werden. Mit jedem Auswahlvorgang werden Ihrer Datendatei eine oder mehrere neue Variablen hinzugefügt.

Vorhergesagte Werte. Dies sind die nach dem Regressionsmodell für jeden Fall vorhergesagten Werte.

- **Nicht standardisiert.** Der Wert, den das Modell für die abhängige Variable vorhersagt.

- **Standardisiert.** Eine Transformation jedes geschätzten Wertes in dessen standardisierte Form. Das heißt, daß die Differenz zwischen dem vorhergesagten Wert und dem mittleren vorhergesagten Wert durch die Standardabweichung der vorhergesagten Werte geteilt wird. Standardisierte, geschätzte Werte haben einen Mittelwert von 0 und eine Standardabweichung von 1.
- **Korrigiert.** Der vorhergesagte Wert für einen Fall, wenn dieser Fall von der Berechnung der Regressionskoeffizienten ausgeschlossen ist.
- **Standardfehler des Mittelwerts.** Standardfehler der vorhergesagten Werte. Ein Schätzwert der Standardabweichung des Durchschnittswertes der abhängigen Variablen für die Fälle, die dieselben Werte für die unabhängigen Variablen haben.

Distanzen. Dies sind Maße zum Auffinden von Fällen mit ungewöhnlichen Wertekombinationen bei der unabhängigen Variablen und von Fällen, die einen großen Einfluß auf das Modell haben könnten.

- **Mahalanobis.** Dieses Maß legt fest, wie weit die Werte der unabhängigen Variablen eines Falles vom Mittelwert aller Fälle abweichen. Eine große Mahalanobis-Abstand charakterisiert einen Fall, der bei einer oder mehreren unabhängigen Variablen Extremwerte besitzt.
- **Cook.** Dies ist ein Maß dafür, wie stark sich die Residuen aller Fälle ändern würden, wenn ein spezieller Fall von der Berechnung der Regressionskoeffizienten ausgeschlossen würde. Ein großer Wert der Cook-Distanz zeigt an, daß der Ausschluß eines Falles von der Berechnung der Regressionskoeffizienten die Koeffizienten substantiell verändert.
- **Hebelwerte.** Werte, die den Einfluß eines Punktes auf die Anpassung der Regression messen. Der zentrierte Wert für die Hebelwirkung bewegt sich zwischen 0 (kein Einfluß auf die Anpassung) und (N-1)/N.

Vorhersageintervalle. Die oberen und unteren Grenzen sowohl für Mittelwert als auch für einzelne Vorhersageintervalle.

- **Mittelwert (Confidence Interval).** Unter- und Obergrenze (zwei Variablen) für das Vorhersageintervall für den mittleren vorhergesagten Wert.

- **Individuell.** Untere und obere Grenzen (zwei Variablen) für das Vorhersageintervall einer abhängigen Variablen für einen einzelnen Fall.
- **Konfidenzintervall.** Geben Sie einen Wert zwischen 1 und 99,99 ein, um das Konfidenzniveau für die beiden Vorhersageintervalle festzulegen. Wählen Sie "Mittelwert" oder "Individuell" aus, bevor Sie diesen Wert eingeben. Typische Werte für Konfidenzniveaus sind 90, 95 und 99.

Residuen. Der tatsächliche Wert der abhängigen Variablen minus dem vorhergesagten Wert aus der Regressionsgleichung.

- **Nicht standardisiert.** Die Differenz zwischen einem beobachteten Wert und dem durch das Modell vorhergesagten Wert.
- **Standardisiert.** Der Quotient aus dem Residuum und einem Schätzer seines Standardfehlers. Standardisierte Residuen, auch bekannt als Pearson-Residuen, haben einen Mittelwert von 0 und eine Standardabweichung von 1.
- **Studentisiert.** Ein Residuum, das durch seine geschätzte Standardabweichung geteilt wird, die je nach der Distanz zwischen den Werten der unabhängigen Variablen des Falles und dem Mittelwert der unabhängigen Variablen von Fall zu Fall variiert.
- **Ausgeschlossen.** Das Residuum für einen Fall, wenn dieser Fall nicht in die Berechnung der Regressionskoeffizienten eingegangen ist. Es ist die Differenz zwischen dem Wert der abhängigen Variablen und dem korrigierten Schätzwert.
- **Studentisiert, ausgeschl..** Der Quotient aus dem ausgeschlossenen Residuum eines Falles und seinem Standardfehler. Die Differenz zwischen einem studentisierten, ausgeschlossenen Residuum und dem entsprechenden studentisierten Residuum gibt an, welchen Unterschied die Entfernung eines Falles für dessen eigene Vorhersage bewirkt.

Einflußstatistiken. Die Änderung in den Regressionskoeffizienten (DfBeta(s)) und vorhergesagten Werten (DfFit), die sich aus dem Ausschluß eines bestimmten Falls ergibt. Standardisierte DfBetas- und DfFit-Werte stehen zusammen mit dem Kovarianzverhältnis zur Verfügung.

- **Differenz in Beta.** Die Differenz im Beta-Wert entspricht der Änderung im Regressionskoeffizienten, die sich aus dem Ausschluß eines bestimmten Falls ergibt. Für jeden Term im Modell, einschließlich der Konstanten, wird ein Wert berechnet.

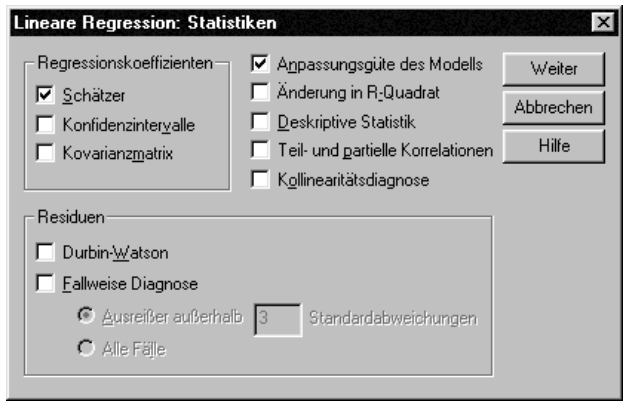
- **Standardisiertes DfBeta.** Die standardisierte Differenz im Beta-Wert. Die Änderung des Regressionskoeffizienten, die sich durch den Ausschluß eines bestimmten Falls ergibt. Es empfiehlt sich, Fälle mit absoluten Werten größer als 2 geteilt durch die Quadratwurzel von N zu überprüfen, wenn N die Anzahl der Fälle darstellt. Für jeden Term im Modell, einschließlich der Konstanten, wird ein Wert berechnet,
- **Differenz im vorhergesagten Wert.** Die Differenz im vorhergesagten Wert ist die Änderung im vorhergesagten Wert, die sich aus dem Ausschluß eines bestimmten Falls ergibt.
- **Standardisiertes DfFit.** Die standardisierte Differenz im Anpassungswert. Die Änderung des Schätzwertes, die sich durch Ausschluß eines bestimmten Falls ergibt. Es empfiehlt sich, Fälle mit absoluten Werten größer als 2 geteilt durch die Quadratwurzel von p/N zu überprüfen, wobei p die Anzahl der unabhängigen Variablen in der Gleichung und N die Anzahl der Fälle darstellt.
- **Kovarianzverhältnis.** Das Verhältnis der Determinante der Kovarianzmatrix bei Ausschluß eines bestimmten Falles von der Berechnung des Regressionskoeffizienten zur Determinante der Kovarianzmatrix bei Einschluß aller Fälle. Wenn der Quotient dicht bei 1 liegt, beeinflußt der ausgeschlossene Fall die Kovarianzmatrix nur unwesentlich.

Koeffizientenstatistik. Speichert den Regressionskoeffizienten in einem Daten-Set oder in einer Datendatei. Daten-Sets sind für die anschließende Verwendung in der gleichen Sitzung verfügbar, werden jedoch nicht als Dateien gespeichert, sofern Sie diese nicht ausdrücklich vor dem Beenden der Sitzung speichern. Die Namen von Daten-Sets müssen den SPSS-Regeln zum Benennen von Variablen entsprechen. Für weitere Informationen siehe “Variablennamen” in Kapitel 5 auf S. 99.

Modellinformation in XML-Datei exportieren. Parameterschätzer und (wahlweise) ihre Kovarianzen werden in die angegebene Datei exportiert. SmartScore und SPSS Server (gesondertes Produkt) können anhand dieser Modelldatei die Modellinformationen zu Bewertungszwecken auf andere Datendateien anwenden.

Lineare Regression: Statistiken

Abbildung 27-6
Dialogfeld "Lineare Regression: Statistiken"



Folgende Statistiken sind verfügbar:

Regressionskoeffizienten. Mit Schätzer zeigen Sie den Regressionskoeffizienten *B*, den Standardfehler von *B*, das Beta des standardisierten Koeffizienten, den *t*-Wert für *B* und das zweiseitige Signifikanzniveau von *t* an. Mit Konfidenzintervalle werden die 95%-Konfidenzintervalle für jeden Regressionskoeffizienten oder eine Kovarianzmatrix angezeigt. Mit Kovarianzmatrix wird eine Varianz-Kovarianz-Matrix von Regressionskoeffizienten mit Kovarianzen angezeigt, die nicht auf der Diagonalen liegen, und Varianzen, die auf der Diagonalen liegen. Außerdem wird eine Korrelationsmatrix angezeigt.

Anpassungsgüte des Modells. Die aufgenommenen und entfernten Variablen aus dem Modell werden aufgelistet, und die folgenden Statistiken der Anpassungsgüte werden angezeigt: multiples *R*, R^2 und korrigiertes R^2, Standardfehler der Differenz und eine Tabelle zur Varianzanalyse.

Änderung in R-Quadrat. Die Änderung in R^2, die aus dem Hinzufügen oder Entfernen einer unabhängigen Variablen resultiert. Wenn die durch eine Variable bewirkte Änderung in R^2 groß ist, bedeutet dies, daß diese Variable eine aussagekräftige Einflußvariable für die abhängige Variable ist.

Deskriptive Statistiken. Liefert die Anzahl gültiger Fälle, Mittelwert und Standardabweichung für jede Variable in der Analyse. Außerdem werden eine Korrelationsmatrix mit einem einseitigen Signifikanzniveau und die Anzahl der Fälle für jede Korrelation angezeigt.

Partielle Korrelation. Die Korrelation, die zwischen zwei Variablen verbleibt, nachdem die Korrelation entfernt wurde, die aus dem wechselseitigen Zusammenhang mit den anderen Variablen stammt. Die Korrelation zwischen der abhängigen Variablen und einer unabhängigen Variablen, wenn die linearen Effekte der anderen unabhängigen Variablen im Modell aus beiden entfernt wurden.

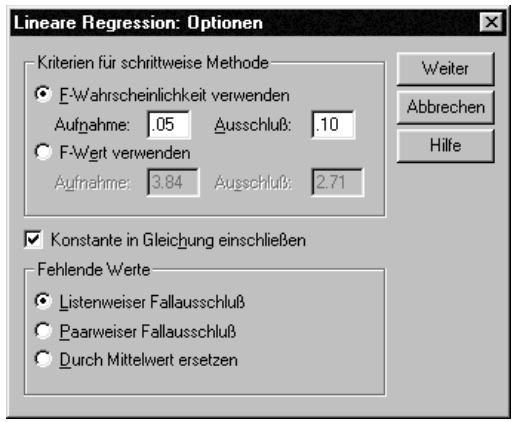
Teilkorrelation (Pivot Table Regression). Die Korrelation zwischen der abhängigen Variablen und einer unabhängigen Variablen, wenn die linearen Effekte der anderen unabhängigen Variablen im Modell aus der unabhängigen Variablen entfernt wurden. Die Korrelation entspricht der Änderung in R-Quadrat beim Addieren einer Variable zu einer Gleichung.

Kollinearitätsdiagnose. Kollinearität (oder Multikollinearität) ist die unerwünschte Situation, in der eine unabhängige Variable eine lineare Funktion anderer unabhängiger Variablen ist. Eigenwerte der skalierten und unzentrierten Kreuzproduktmatrix, Bedingungsindexe und Proportionen der Varianzzerlegung werden zusammen mit Varianzfaktoren (VIF) und Toleranzen für einzelne Variablen angezeigt.

Residuen. Hiermit werden der Durbin-Watson-Test für Reihenkorrelationen der Residuen sowie die fallweise Diagnose für die Fälle angezeigt, die das Auswahlkriterium (Ausreißer über *n* Standardabweichungen) erfüllen.

Lineare Regression: Optionen

Abbildung 27-7
Dialogfeld "Lineare Regression: Optionen"



Die folgenden Optionen sind verfügbar:

Kriterien für schrittweise Methode. Diese Optionen eignen sich für den Fall, daß die Vorwärts-, Rückwärts- oder schrittweise Methode der Variablenauswahl angegeben wurde. Variablen im Modell können abhängig entweder von der Signifikanz (Wahrscheinlichkeit) des *F*-Werts oder vom *F*-Wert selbst eingeschlossen oder entfernt werden.

- **F-Wahrscheinlichkeit verwenden.** Eine Variable wird in ein Modell aufgenommen, wenn das Signifikanzniveau ihres F-Wertes kleiner ist als der Aufnahmewert ist. Sie wird ausgeschlossen, wenn das Signifikanzniveau größer ist als der Ausschlußwert. Der Aufnahmewert muß kleiner sein als der Ausschlußwert, und beide Werte müssen positiv sein. Um mehr Variablen in das Modell aufzunehmen, erhöhen Sie den Aufnahmewert. Um mehr Variablen aus dem Modell auszuschließen, senken Sie den Ausschlußwert.
- **F-Wert verwenden.** Eine Variable wird in ein Modell aufgenommen, wenn ihr F-Wert größer als der Aufnahmewert ist. Sie wird ausgeschlossen, wenn der F-Wert kleiner ist als der Ausschlußwert. Der Aufnahmewert muß größer sein als der Ausschlußwert, und beide Werte müssen positiv sein. Um mehr Variablen in das Modell aufzunehmen, senken Sie den Aufnahmewert. Um mehr Variablen aus dem Modell auszuschließen, erhöhen Sie den Ausschlußwert.

Konstante in Gleichung einschließen. Als Voreinstellung enthält das Regressionsmodell einen konstanten Term. Wenn diese Option deaktiviert ist, wird die Regression durch den Ursprung gezwungen (selten verwendet). Manche Resultate einer durch den Ursprung verlaufenden Regression lassen sich nicht mit denen einer Regression vergleichen, die eine Konstante aufweist. Beispielsweise kann R^2 nicht in der üblichen Weise interpretiert werden.

Fehlende Werte. Sie können eine der folgenden Optionen auswählen:

- **Listenweiser Fallausschluß.** Nur Fälle mit gültigen Werten für alle Variablen werden in die Analyse einbezogen.
- **Paarweiser Fallausschluß.** Fälle mit vollständigen Daten für das korrelierte Variablenpaar werden zum Berechnen des Korrelationskoeffizienten verwendet, auf dem die Regressionsanalyse basiert. Freiheitsgrade basieren auf dem minimalen paarweisen *N*.
- **Durch Mittelwert ersetzen.** Alle Fälle werden für Berechnungen verwendet, wobei der Mittelwert der Variablen die fehlenden Beobachtungen ersetzt.

Zusätzliche Funktionen beim Befehl REGRESSION

Mit der Befehlssprache verfügen Sie außerdem über folgende Möglichkeiten:

- Schreiben einer Korrelationsmatrix, oder Einlesen einer Matrix anstelle der Rohdaten, um eine Regressionsanalyse zu erhalten (mit dem Unterbefehl `MATRIX`)
- Angeben von Toleranzniveaus (mit dem Unterbefehl `CRITERIA`)
- Berechnen mehrerer Modelle für dieselben oder unterschiedliche abhängige Variablen (mit den Unterbefehlen `METHOD` und `DEPENDENT`)
- Berechnen zusätzlicher Statistiken (mit den Unterbefehlen `DESCRIPTIVES` und `STATISTICS`)

Vollständige Informationen zur Syntax finden Sie im *SPSS Syntax Reference Guide*.

Kapitel
28

Kurvenanpassung

Mit der Prozedur "Kurvenanpassung" werden Regressionsstatistiken zur Kurvenanpassung und zugehörige Diagramme für 11 verschiedene Regressionsmodelle zur Kurvenanpassung erstellt. Für jede abhängige Variable wird ein separates Modell erstellt. Außerdem können Sie vorhergesagte Werte, Residuen und Vorhersageintervalle als neue Variablen speichern.

Beispiel. Ein Internet-Dienstanbieter verfolgt den Prozentsatz des mit Viren infizierten E-Mail-Verkehrs über die Netzwerke im Lauf der Zeit. Ein Streudiagramm zeigt, daß eine nichtlineare Beziehung vorliegt. Sie können ein quadratisches oder kubisches Modell an die Daten anpassen und die Gültigkeit der Annahmen sowie die Güte der Anpassung des Modells prüfen.

Statistiken. Für jedes Modell: Regressionskoeffizienten, multiples R, R^2, korrigiertes R^2, Standardfehler des Schätzers, Tabelle für die Varianzanalyse, vorhergesagte Werte, Residuen und Vorhersageintervalle. Modelle: linear, logarithmisch, invers, quadratisch, kubisch, Potenz, zusammengesetzt, S-Kurve, logistisch, Wachstum und exponentiell.

Daten. Die abhängigen und die unabhängigen Variablen müssen quantitativ sein. Wenn Sie aus der Arbeitsdatei Zeit als unabhängige Variable ausgewählt haben (statt eine Variable auszuwählen), erzeugt die Prozedur "Kurvenanpassung" eine Zeitvariable mit gleichen Zeitabständen zwischen den Fällen. Wenn Zeit ausgewählt wurde, sollte die abhängige Variable eine Zeitreihenmessung sein. Zur Zeitreihenanalyse ist eine Datendateistruktur erforderlich, in der jeder Fall (jede Zeile) einen Satz von Beobachtungen zu unterschiedlichen Zeiten bei gleichen Zeitabständen zwischen den Fällen darstellt.

Annahmen. Stellen Sie Ihre Daten grafisch dar, um den Zusammenhang zwischen den unabhängigen und den abhängigen Variablen (linear, exponentiell usw.) erkennen zu können. Die Residuen eines guten Modells müssen willkürlich und normalverteilt sein. Bei einem linearen Modell müssen folgende Annahmen erfüllt werden: Für

jeden Wert der unabhängigen Variablen muß die abhängige Variable normalverteilt vorliegen. Die Varianz der Verteilung der abhängigen Variablen muß für alle Werte der unabhängigen Variablen konstant sein. Die abhängige Variable und die unabhängige Variable müssen linear zusammenhängen, und alle Beobachtungen müssen unabhängig sein.

Abbildung 28-1
Zusammenfassungstabelle "Kurvenanpassung"

R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
,885	,783	,766	,083

Die unabhängige Variable ist Hours since detection.

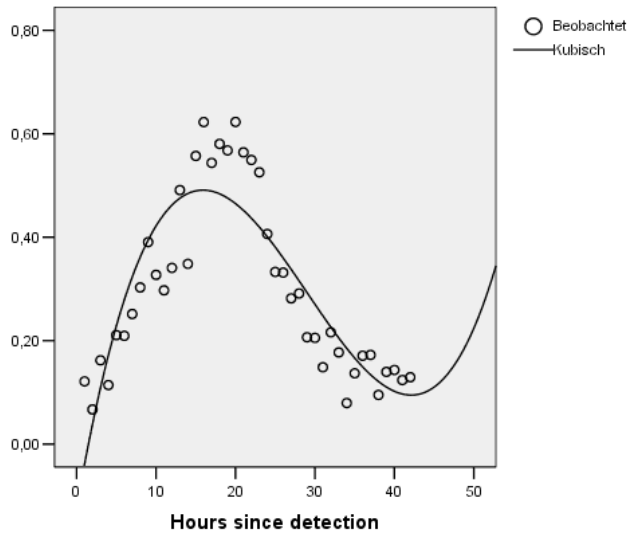
Abbildung 28-2
Kurvenanpassungs-ANOVA

	Quadratsumme	Freiheitsgrade	Mittel der Quadrate	F	Sig.
Regression	,949	3	,316	45,736	,000
Residuen	,263	38	,007		
Gesamt	1,212	41			

Abbildung 28-3
Kurvenanpassungskoeffizienten

	Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten		
	B	Standardfehler	Beta	t	Sig.
Hours since detection	,088	,011	6,305	7,888	,000
Hours since detection ** 2	-,004	,001	-12,111	-6,363	,000
Hours since detection ** 3	4,40E-005	,000	5,638	4,780	,000
(Konstante)	-,123	,056		-2,187	,035

Abbildung 28-4
Kurvenanpassungsdiagramm

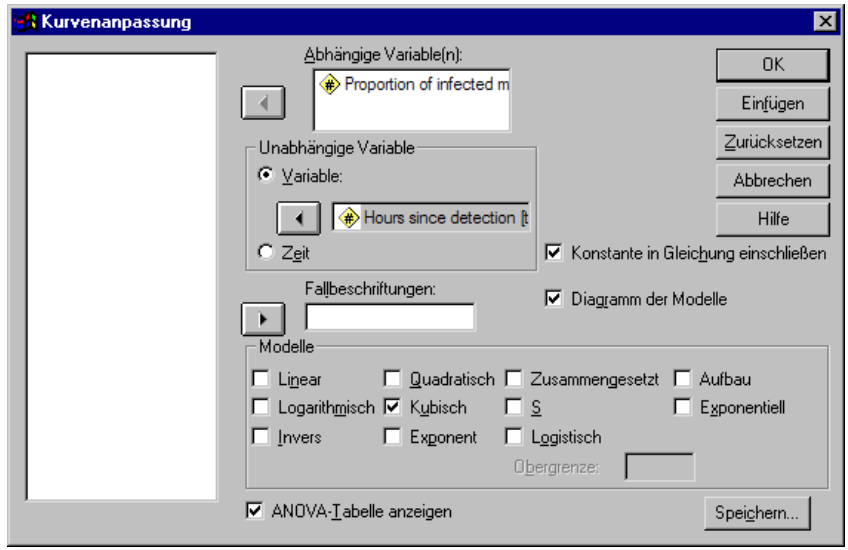


So führen Sie eine Kurvenanpassung durch:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Regression
    Kurvenanpassung…

Abbildung 28-5
Dialogfeld "Kurvenanpassung"



- Wählen Sie eine oder mehrere abhängige Variablen aus. Für jede abhängige Variable wird ein separates Modell erstellt.

- Wählen Sie eine unabhängige Variable aus (wählen Sie entweder eine Variable aus der Arbeitsdatei, oder wählen Sie Zeit aus).

- Die folgenden Optionen sind verfügbar:
  - Eine Variable zum Beschriften der Fälle in Streudiagrammen auswählen. Sie können für jeden Punkt im Streudiagramm das Symbol zum Identifizieren von Punkten verwenden, um den Wert der Variablen für die "Fallbeschriftung" anzeigen zu lassen.
  - Klicken Sie auf Speichern, um vorhergesagte Werte, Residuen und Vorhersageintervalle als neue Variablen zu speichern.

  Außerdem sind folgende Optionen verfügbar:

  - **Konstante in Gleichung einschließen.** Mit dieser Option wird ein konstanter Term in der Regressionsgleichung geschätzt. In der Standardeinstellung ist die Konstante eingeschlossen.

- **Diagramm der Modelle.** Mit dieser Option werden für alle ausgewählten Modelle die Werte der abhängigen Variablen über der unabhängigen Variablen grafisch dargestellt. Für jede abhängige Variable wird ein eigenes Diagramm erzeugt.
- **ANOVA-Tabelle anzeigen.** Mit dieser Option wird für jedes ausgewählte Modell eine Zusammenfassung für die Varianzanalyse angezeigt.

Modelle für die Kurvenanpassung

Sie können ein oder mehrere Regressionsmodelle für die Kurvenanpassung auswählen. Stellen Sie Ihre Daten grafisch dar, um zu ermitteln, welches Modell Sie verwenden sollten. Wenn Ihre Variablen in einem linearen Zusammenhang zu stehen scheinen, verwenden Sie ein einfaches lineares Regressionsmodell. Wenn Ihre Variablen in keinem linearen Zusammenhang stehen, transformieren Sie diese. Wenn eine Transformation keine Abhilfe schafft, benötigen Sie möglicherweise ein komplizierteres Modell. Betrachten Sie ein Streudiagramm Ihrer Daten. Wenn das Diagramm einer Ihnen bekannten mathematischen Funktion ähnelt, passen Sie Ihre Daten an diesen Modelltyp an. Wenn Ihre Daten zum Beispiel einer Exponentialfunktion ähneln, verwenden Sie ein exponentielles Modell.

Linear. Die Gleichung für dieses Modell lautet Y = b0 + (b1 * t). Die Werte der Zeitreihe werden als lineare Funktion der Zeit aufgefaßt.

Logarithmisch. Die Gleichung für dieses Modell lautet Y = b0 + (b1 * ln(t)).

Invers. Ein Modell mit der Gleichung Y = b0 + (b1 / t).

Quadratisch. Die Gleichung für dieses Modell lautet: Y = b0 + (b1 * t) + (b2 * t**2). Das quadratische Modell kann zum Modellieren von Zeitreihen verwendet werden, die "abheben" oder gedämpft verlaufen.

Kubisch. Ein Modell mit folgender Gleichung:Y = b0 + (b1 * t) + (b2 * t**2) + (b3 * t**3).

Exponent. Die Gleichung für dieses Modell lautet: Y = b0 * (t**b1) oder ln(Y) = ln(b0) + (b1 * ln(t)).

Zusammengesetzt. Dieses Modell basiert auf folgender Gleichung: Y = b0 * (b1**t) oder ln(Y) = ln(b0) + (ln(b1) * t).

S-Kurve. Ein Modell, dessen Gleichung lautet: Y = e**(b0 + (b1/t)) oder ln(Y) = b0 + (b1/t).

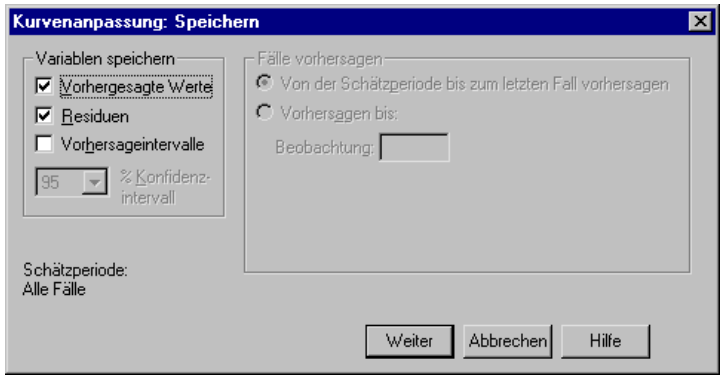
Logistisch. Die Gleichung für dieses Modell lautet Y = 1 / (1/u + (b0 * (b1**t))) oder ln(1/y-1/u) = ln (b0) + (ln(b1) * t), wobei u die obere Schranke ist. Nach der Auswahl von "Logistisch" muss der Wert der oberen Schranke angegeben werden, der in der Regressionsgleichung verwendet werden soll. Der Wert muss eine positive Zahl sein, die größer ist als der größte Wert der abhängigen Variablen.

Wachstum. Die Gleichung für dieses Modell lautet: Y = e**(b0 + (b1 * t)) oder ln (Y) = b0 + (b1 * t).

Exponentiell. Ein Modell mit folgender Gleichung: Y = b0 * (e**(b1 * t)) oder ln (Y) = ln (b0) + (b1 * t).

Kurvenanpassung: Speichern

Abbildung 28-6
Dialogfeld "Kurvenanpassung: Speichern"



Variablen speichern. Für jedes ausgewählte Modell können Sie vorhergesagte Werte, Residuen (beobachteter Wert der abhängigen Variablen minus vorhergesagter Wert des Modells) und Vorhersageintervalle (Ober- und Untergrenzen) speichern. Die neuen Variablennamen werden mit den beschreibenden Labels in einer Tabelle im Ausgabefenster angezeigt.

Fälle vorhersagen. Wenn Sie in der Arbeitsdatei statt einer Variablen Zeit als unabhängige Variable ausgewählt haben, können Sie nach dem Ende der Zeitreihe eine Vorhersageperiode angeben. Sie können eine der folgenden Möglichkeiten wählen:

- **Von der Schätzperiode bis zum letzten Fall vorhersagen.** Hiermit werden auf der Grundlage der Fälle in der Schätzperiode Werte für alle Fälle in der Datei vorhergesagt. Die unten im Dialogfeld angezeigte Schätzperiode wird im Menü

"Daten", Option "Fälle auswählen", Dialogfeld "Fälle auswählen:Bereich" festgelegt. Wenn keine Schätzperiode definiert wurde, werden alle Fälle zum Schätzen der Werte verwendet.

- **Vorhersagen bis.** Hiermit werden auf der Grundlage der Fälle in der Schätzperiode Werte bis zum angegebenen Datum, zur angegebenen Uhrzeit oder zur angegebenen Beobachtungsnummer vorhergesagt. Mit dieser Funktion können Werte nach dem letzten Fall in der Zeitreihe vorhergesagt werden. Die gegenwärtig definierten Datumsvariablen bestimmen, welche Textfelder zur Verfügung stehen, um das Ende der Vorhersageperiode anzugeben. Wenn keine Datumsvariablen definiert sind, können Sie die letzte Beobachtungs- bzw. Fallnummer angeben.

Datumsvariablen erstellen Sie im Menü "Daten" mit der Option "Datum definieren".

Kapitel
29

Diskriminanzanalyse

Die Diskriminanzanalyse ist nützlich, um ein Vorhersagemodell der Gruppenzugehörigkeit zu entwickeln, das auf den beobachteten Eigenschaften der einzelnen Fälle beruht. Diese Prozedur erzeugt eine Diskriminanzfunktion (oder, bei mehr als zwei Gruppen, ein Set von Diskriminanzfunktionen) auf der Grundlage derjenigen linearen Kombinationen der Einflußvariablen, welche die beste Diskriminanz zwischen den Gruppen ergeben. Die Funktionen werden aus einer Stichprobe der Fälle erzeugt, bei denen die Gruppenzugehörigkeit bekannt ist. Diese Funktionen können dann auf neue Fälle mit Messungen für die Einflußvariablen, aber unbekannter Gruppenzugehörigkeit angewandt werden.

Anmerkung: Die Gruppenvariable kann mehr als zwei Werte besitzen. Die Codes für die Gruppenvariable müssen allerdings ganzzahlige Werte sein, und Sie müssen hierfür die minimalen und maximalen Werte festlegen. Fälle mit Werten außerhalb dieser Grenzen werden von der Analyse ausgeschlossen.

Beispiel. Im Durchschnitt verbrauchen Personen in kühlen Ländern mehr Kalorien pro Tag als Bewohner der Tropen, und ein größerer Anteil der Personen in den kühlen Ländern sind Stadtbewohner. Ein Forscher möchte diese Informationen in einer Funktion zusammenfassen, um zu bestimmen, wie gut eine bestimmte Person diesen beiden Ländergruppen zugeordnet werden kann. Der Forscher nimmt an, daß auch die Bevölkerungsgröße und Wirtschaftsinformationen relevant sein könnten. Mit der Diskriminanzanalyse können Sie die Koeffizienten der linearen Diskriminanzfunktion schätzen, die im Prinzip genauso wie die rechte Seite einer Regressionsgleichung bei mehrfacher Regression aufgebaut ist. Unter Verwendung der Koeffizienten *a*, *b*, *c* und *d* lautet die Funktion also:

D = a * Klima + b * Städtisch + c * Bevölkerung + d * Bruttosozialprodukt der Region je Einwohner.

Wenn diese Variablen für die Unterscheidung zwischen den beiden Klimazonen relevant sind, müssen sich die Werte von *D* für tropische und kühlere Länder unterscheiden. Falls Sie eine schrittweise Methode für die Variablenauswahl verwenden, stellen Sie unter Umständen fest, daß nicht alle vier Variablen in die Funktion aufgenommen werden müssen.

Statistiken. Für jede Variable: Mittelwerte, Standardabweichungen, univariate ANOVA. Für jede Analyse: Box-*M*, Korrelationsmatrix innerhalb der Gruppen, Kovarianzmatrix innerhalb der Gruppen, Kovarianzmatrix der einzelnen Gruppen, gesamte Kovarianzmatrix. Für jede kanonische Diskriminanzfunktion: Eigenwert, Prozentwert der Varianz, kanonische Korrelation, Wilks-Lambda, Chi-Quadrat. Für jeden Schritt: a-priori-Wahrscheinlichkeit, Funktionskoeffizienten nach Fisher, nicht standardisierte Funktionskoeffizienten, Wilks-Lambda für jede kanonische Funktion.

Daten. Die Gruppenvariable muß über eine begrenzte Anzahl unterschiedener Kategorien verfügen, die als ganzzahlige Werte kodiert werden. Unabhängige nominale Variablen müssen in Dummy- oder Kontrastvariablen umkodiert werden.

Annahmen. Die Fälle müssen unabhängig sein. Einflußvariablen müssen in multivariater Normalverteilung vorliegen, und die Varianz-Kovarianz-Matrizen innerhalb der Gruppen müssen zwischen den Gruppen gleich groß sein. Die Gruppenzugehörigkeit muß sich wechselseitig ausschließen (das heißt, kein Fall gehört zu mehr als einer Gruppe) und umfassend sein (das heißt, alle Fälle gehören zu einer Gruppe). Diese Prozedur ist am effektivsten, wenn die Gruppenzugehörigkeit eine rein kategoriale Variable ist. Wenn die Gruppenzugehörigkeit hingegen auf den Werten einer stetigen Variablen basiert (zum Beispiel bei einem Vergleich von IQ-Werten), sollten Sie die lineare Regression in Betracht ziehen, um von den reichhaltigeren Informationen zu profitieren, die in der stetigen Variablen selbst enthalten sind.

Abbildung 29-1
Ausgabe der Diskriminanzanalyse

Eigenwerte

		Eigenwert	% der Varianz	Kumulierte %	Kanonische Korrelation
Funktion	1	1.002	100.0	100.0	.707

Wilks' Lambda

Test der Funktion(en)		Wilks-Lambda	Chi-Quadrat	df	Signifikanz
	1	.499	31.934	4	.000

Funktionen bei den Gruppen-Zentroiden

		Funktion
		1
KLIMA		
	Tropisch	-.869
	Temperiert	1.107

Struktur-Matrix

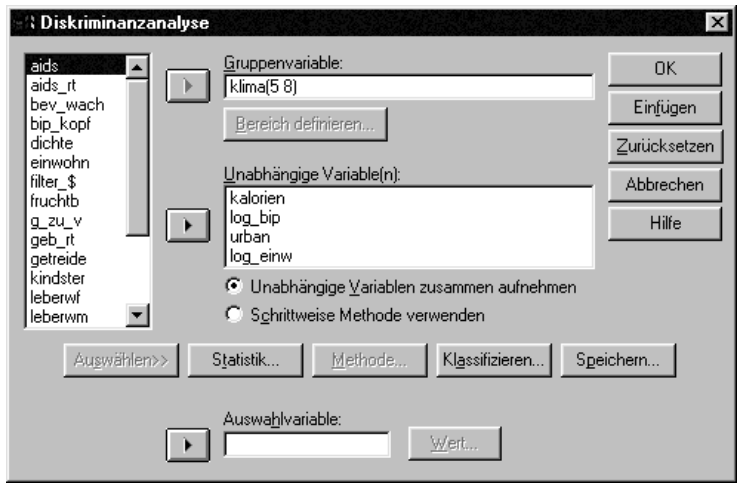
	Funktion
	1
CALORIES	.986
LOG_GDP	.790
URBAN	.488
LOG_POP	.082

So lassen Sie eine Diskriminanzanalyse berechnen

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Klassifizieren
  Diskriminanzanalyse...

Abbildung 29-2
Dialogfeld "Diskriminanzanalyse"

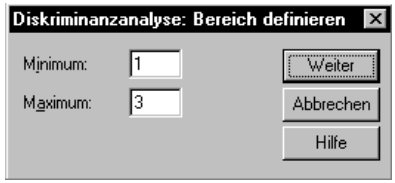


- Wählen Sie eine Gruppenvariable mit ganzzahligen Werten aus, und klicken Sie auf Bereich definieren, um die gewünschten Kategorien festzulegen.

- Wählen Sie die unabhängigen Variablen oder Einflußvariablen aus. (Wenn die Gruppenvariable nicht ganzzahlig ist, können Sie eine Variable mit dieser Eigenschaft im Menü "Transformieren" mit dem Befehl "Automatisch umkodieren" erstellen.)
- Wählen Sie die gewünschte Methode für die Eingabe der unabhängigen Variablen aus.
  - **Unabhängige Variablen zusammen aufnehmen.** Nimmt alle unabhängigen Variablen, welche die Toleranzkriterien erfülllen, gleichzeitig auf.
  - **Schrittweise Methode verwenden.** Verwendet ein schrittweises Verfahren zur Steuerung von Variablenaufnahme und Variablenausschluß.
- Wahlweise können Sie die Fälle auch mit Hilfe einer Auswahlvariablen auswählen.

Diskriminanzanalyse: Bereich definieren

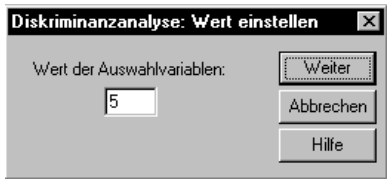
Abbildung 29-3
Dialogfeld "Diskriminanzanalyse: Bereich definieren"



Geben Sie den kleinsten (Minimum) und den größten (Maximum) Wert der Gruppenvariablen für die Analyse an. Fälle mit Werten außerhalb dieses Bereichs werden in der Diskriminanzanalyse nicht verwendet, aber ausgehend von den Ergebnissen der Analyse in eine der vorhandenen Gruppen eingeordnet. Die Minimum- und Maximumwerte müssen ganzzahlig sein.

Diskriminanzanalyse: Fälle auswählen

Abbildung 29-4
Dialogfeld "Diskriminanzanalyse: Wert einstellen"



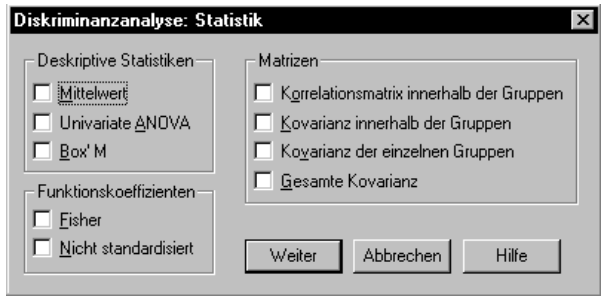
So wählen Sie die Fälle für die Analyse aus:

- Wählen Sie im Dialogfeld "Diskriminanzanalyse" eine Auswahlvariable aus.
- Klicken Sie auf Wert, um eine Ganzzahl als Auswahlvariable einzugeben.

Bei der Ableitung der Diskriminanzfunktionen werden nur die Fälle verwendet, deren Auswahlvariablen den angegebenen Wert aufweisen. Statistiken und Klassifikationsergebnisse werden sowohl für die ausgewählten als auch für die nicht ausgewählten Fälle erzeugt. Mit diesem Prozess liegt ein Mechanismus vor, mit dem neue Fälle anhand von bereits vorhandenen Daten klassifiziert werden können, oder mit dem Sie Ihre Daten in Teilmengen von Lern- und Testfällen einteilen können, um so eine Gültigkeitsprüfung des erzeugten Modells durchzuführen.

Diskriminanzanalyse: Statistik

Abbildung 29-5
Dialogfeld "Diskriminanzanalyse: Statistik"



Deskriptive Statistiken. Verfügbare Optionen sind Mittelwerte (einschließlich Standardabweichungen), univariate ANOVA und der Box-*M*-Test.

- **Mittelwerte (Discriminant Analysis).** Zeigt Gesamt- und Gruppenmittelwerte sowie Standardabweichungen für die unabhängigen Variablen an.
- **Univariate ANOVA (Discriminant Analysis).** Führt für jede unabhängige Variable eine einfaktorielle Varianzanalyse durch, d. h. einen Test auf Gleichheit der Gruppenmittelwerte .
- **Box-M.** Ein Test auf Gleichheit der Kovarianzmatrizen der Gruppen. Bei hinreichend großen Stichproben bedeutet ein nichtsignifikanter p-Wert, daß die Anhaltspunkte für unterschiedliche Kovarianzmatrizen nicht hinreichend

sind. Der Test ist empfindlich gegenüber Abweichungen von der multivariaten Normalverteilung.

Funktionskoeffizienten. Verfügbare Optionen sind Klassifikationskoeffizienten nach Fisher und nicht standardisierte Koeffizienten.

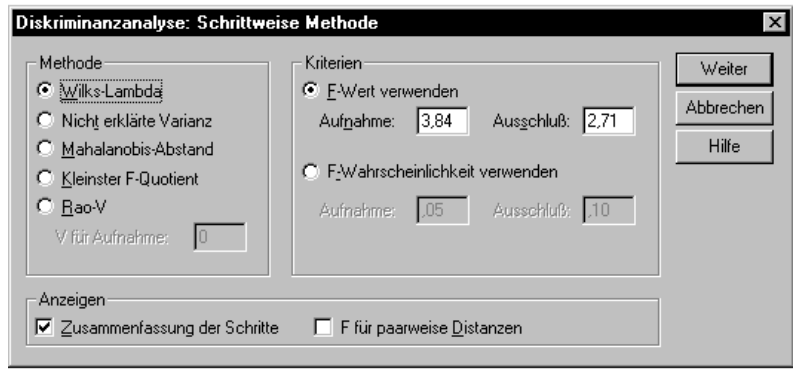
- **Klassifizierungsfunktion nach Fisher (Discriminant Analysis).** Zeigt die Koeffizienten der Klassifizierungsfunktion nach Fisher an, die direkt für die Klassifizierung verwendet werden können.Es wird ein Satz von Koeffizienten für jede Gruppe ermittelt. Ein Fall wird der Gruppe zugewiesen, für den er den größten Diskriminanzwert aufweist.
- **Nicht standardisiert (Discriminant Analysis).** Zeigt die nicht standardisierten Koeffizienten der Diskriminanzfunktion an.

Matrizen. Als Koeffizientenmatrizen für unabhängige Variablen stehen die Korrelationsmatrix innerhalb der Gruppen, die Kovarianzmatrix innerhalb der Gruppen, die Kovarianzmatrix der einzelnen Gruppen und die gesamte Kovarianzmatrix zur Verfügung.

- **Korrelationsmatrix innerhalb der Gruppen.** Zeigt eine gemeinsame Korrelationsmatrix innerhalb der Gruppen an, die als Mittel der separaten Kovarianzmatrizen für alle Gruppen vor der Berechnung der Korrelationen bestimmt wird.
- **Kovarianz innerhalb der Gruppen.** Zeigt eine gemeinsame Kovarianzmatrix innerhalb der Gruppen an, die sich von der Gesamt-Kovarianzmatrix unterscheiden kann. Die Matrix wird als Mittel der einzelnen Kovarianzmatrizen für alle Gruppen berechnet.
- **Kovarianz der einzelnen Gruppen.** Zeigt separate Kovarianzmatrizen für jede Gruppe an.
- **Gesamte Kovarianz.** Zeigt die Kovarianzmatrix für alle Fälle an, so als wären sie aus einer einzigen Stichprobe.

Diskriminanzanalyse: Schrittweise Methode

Abbildung 29-6
Dialogfeld "Diskriminanzanalyse: Schrittweise Methode"



Methode. Wählen Sie die Statistiken aus, die für die Aufnahme oder den Ausschluß neuer Variablen dienen sollen. Die Optionen Wilks-Lambda, nicht erklärte Varianz, Mahalanobis-Abstand, kleinster *F*-Quotient und Rao-*V* stehen zur Verfügung. Mit Rao-*V* können Sie den Mindestanstieg von *V* für eine einzugebende Variable angeben.

- **Wilks-Lambda.** Eine Auswahlmethode für Variablen bei der schrittweisen Diskriminanzanalyse. Die Aufnahme von Variablen in die Gleichung erfolgt anhand der jeweiligen Verringerung von Wilks-Lambda. Bei jedem Schritt wird diejenige Variable aufgenommen, die den Gesamtwert von Wilks-Lambda am meisten vermindert.
- **Nicht erklärte Varianz.** Bei jedem Schritt wird die Variable aufgenommen, welche die Summe der nicht erklärten Streuung zwischen den Gruppen minimiert.
- **Mahalanobis-Abstand.** Dieses Maß legt fest, wie weit die Werte der unabhängigen Variablen eines Falles vom Mittelwert aller Fälle abweichen. Eine große Mahalanobis-Abstand charakterisiert einen Fall, der bei einer oder mehreren unabhängigen Variablen Extremwerte besitzt.
- **Kleinster F-Quotient.** Eine Methode für die Variablenselektion in einer schrittweisen Analyse. Sie beruht auf der Maximierung eines F-Quotienten, der aus dem Mahalanobis-Abstand zwischen den Gruppen errechnet wird.
- **Rao-V.** Ein Maß für die Unterschiede zwischen Gruppenmittelwerten, auch Lawley-Hotelling-Spur genannt. Bei jedem Schritt wird die Variable aufgenommen, die den Anstieg des Rao-V maximiert. Wenn Sie diese Option

ausgewählt haben, geben Sie den Minimalwert ein, den eine Variable für die Aufnahme in die Analyse aufweisen muß.

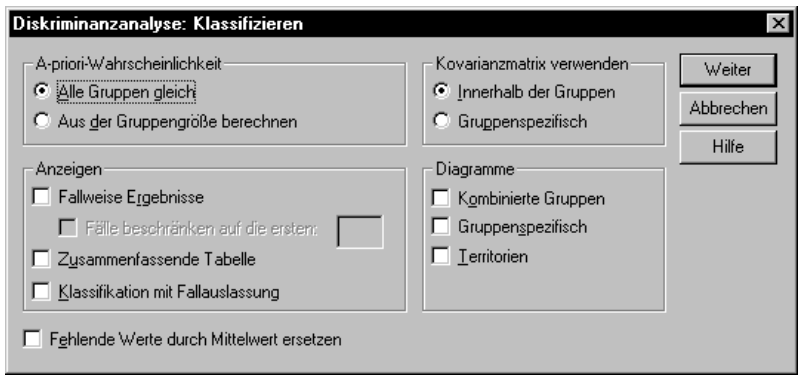
Kriterien. Verfügbar sind F-Wert verwenden und F-Wahrscheinlichkeit verwenden. Geben Sie Werte für die Aufnahme und den Ausschluß der Variablen an.

- **F-Wert verwenden.** Eine Variable wird in ein Modell aufgenommen, wenn ihr F-Wert größer als der Aufnahmewert ist. Sie wird ausgeschlossen, wenn der F-Wert kleiner ist als der Ausschlußwert. Der Aufnahmewert muß größer sein als der Ausschlußwert, und beide Werte müssen positiv sein. Um mehr Variablen in das Modell aufzunehmen, senken Sie den Aufnahmewert. Um mehr Variablen aus dem Modell auszuschließen, erhöhen Sie den Ausschlußwert.
- **F-Wahrscheinlichkeit verwenden.** Eine Variable wird in ein Modell aufgenommen, wenn das Signifikanzniveau ihres F-Wertes kleiner ist als der Aufnahmewert ist. Sie wird ausgeschlossen, wenn das Signifikanzniveau größer ist als der Ausschlußwert. Der Aufnahmewert muß kleiner sein als der Ausschlußwert, und beide Werte müssen positiv sein. Um mehr Variablen in das Modell aufzunehmen, erhöhen Sie den Aufnahmewert. Um mehr Variablen aus dem Modell auszuschließen, senken Sie den Ausschlußwert.

Anzeigen. Mit Zusammenfassung der Schritte können Sie nach jedem Schritt die Statistiken für alle Variablen anzeigen lassen. Bei Auswahl von F für paarweise Distanzen wird für jedes Gruppenpaar eine Matrix des paarweisen *F*-Quotienten angezeigt.

Diskirminanzanalyse: Klassifizieren

Abbildung 29-7
Dialogfeld "Diskriminanzanalyse: Klassifizieren"



A-priori-Wahrscheinlichkeit. Diese Werte werden bei der Klassifikation verwendet. Sie können für alle Gruppen die gleiche a-priori-Wahrscheinlichkeit festlegen (Alle Gruppe gleich), oder Sie können bestimmen, dass die beobachteten Gruppengrößen in der Stichprobe die Wahrscheinlichkeiten der Gruppenzugehörigkeit bestimmen (Aus der Gruppengröße berechnen).

Anzeigen. Die verfügbaren Anzeigeoptionen lauten: "Fallweise Ergebnisse", "Zusammenfassende Tabelle" und "Klassifikation mit Fallauslassung".

- **Fallweise Ergebnisse.** Für jeden Fall werden Codes für die tatsächliche Gruppe, die vorhergesagte Gruppe, a-posteriori-Wahrscheinlichkeiten und Diskriminanzwerte angezeigt.
- **Zusammenfassende Tabelle.** Die Anzahl der Fälle, die auf Grundlage der Diskriminanzanalyse jeder der Gruppen richtig oder falsch zugeordnet werden. Diese Aufstellung wird auch als Klassifikationsmatrix bezeichnet.
- **Klassifikation mit Fallauslassung.** Jeder Fall der Analyse wird durch Funktionen aus allen anderen Fällen unter Auslassung dieses Falls klassifiziert. Diese Klassifikation wird auch als "U-Methode" bezeichnet.

Fehlende Werte durch Mittelwert ersetzen. Wenn Sie diese Option wählen, werden fehlende Werte durch den Mittelwert der jeweiligen unabhängigen Variablen ersetzt, allerdings nur während der Klassifikation der Gruppen.

Kovarianzmatrix verwenden. Sie können wählen, ob zur Klassifikation der Fälle die Kovarianzmatrix innerhalb der Gruppen oder die gruppenspezifische Kovarianzmatrix verwendet werden soll.

- **Innerhalb der Gruppen.** Zur Klassifizierung von Fällen wird die gemeinsame Kovarianzmatrix innerhalb der Gruppen verwendet.
- **Gruppenspezifisch.** Für die Klassifizierung werden gruppenspezifische Kovarianzmatrizen verwendet. Da die Klassifizierung auf Diskriminanzfunktionen und nicht auf ursprünglichen Variablen basiert, entspricht diese Option nicht immer der Verwendung einer quadratischen Diskriminanzfunktion.

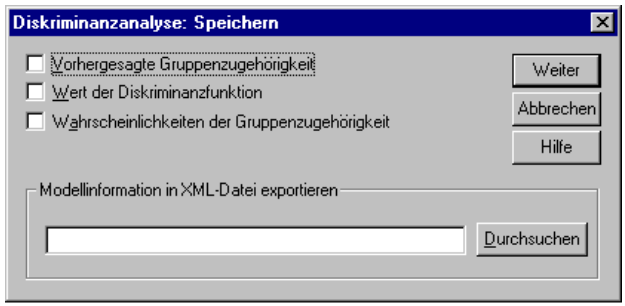
Diagramme. Die verfügbaren Diagrammoptionen sind "Kombinierte Gruppen", "Gruppenspezifisch" und "Territorien".

- **Kombinierte Gruppen.** Erzeugt ein alle Gruppen umfassendes Streudiagramm der Werte für die ersten beiden Diskriminanzfunktionen. Wenn nur eine Funktion vorliegt, wird statt dessen ein Histogramm angezeigt.

- **Gruppenspezifisch.** Erzeugt gruppenspezifische Streudiagramme der Werte für die ersten beiden Diskriminanzfunktionen. Wenn nur eine Funktion vorliegt, wird statt dessen ein Histogramm angezeigt.
- **Territorien.** Ein Diagramm der Grenzen, mit denen Fälle auf der Grundlage von Funktionswerten in Gruppen klassifiziert werden. Die Zahlen entsprechen den Gruppen, in die die Fälle klassifiziert wurden. Der Mittelwert jeder Gruppe wird durch einen darin liegenden Stern (*) angezeigt. Dieses Diagramm wird nicht angezeigt, wenn nur eine Diskriminanzfunktion vorliegt.

Diskriminanzanalyse: Speichern

Abbildung 29-8
Dialogfeld "Diskriminanzanalyse: Speichern"



Sie können der aktiven Datendatei neue Variablen hinzufügen. Die verfügbaren Optionen sind "Vorhergesagte Gruppenzugehörigkeit" (eine einzelne Variable), "Wert der Diskriminanzfunktion" (eine Variable für jede Diskriminanzfunktion in der Lösung) und "Wahrscheinlichkeiten der Gruppenzugehörigkeit" unter Berücksichtigung der Werte der Diskriminanzfunktion (eine Variable pro Gruppe).

Des weiteren können Sie Modellinformationen in die angegebene Datei exportieren. SmartScore und SPSS Server (gesondertes Produkt) können anhand dieser Modelldatei die Modellinformationen zu Bewertungszwecken auf andere Datendateien anwenden.

Zusätzliche Funktionen beim Befehl DISCRIMINANT

Mit der SPSS-Befehlssprache verfügen Sie über die folgenden zusätzlichen Möglichkeiten:

- Durchführen von mehreren Diskriminanzanalysen (mit einem Befehl) und Festlegen der Reihenfolge, in der die Variablen eingegeben werden (mit dem Unterbefehl `ANALYSIS`).
- Eingeben von a-priori-Wahrscheinlichkeiten für den Klassifikation (mit dem Unterbefehl `PRIORS`).
- Anzeigen von rotierten Mustern und Strukturmatrizen (mit dem Unterbefehl `ROTATE`).
- Begrenzen der Anzahl von extrahierten Diskriminanzfunktionen (mit dem Unterbefehl `FUNCTIONS`).
- Beschränken der Klassifikation auf die Fälle, die für die Analyse ausgewählt (oder nicht ausgewählt) wurden (mit dem Unterbefehl `SELECT`).
- Einlesen und Analysieren der Korrelationsmatrix (mit dem Unterbefehl `MATRIX`).
- Schreiben einer Korrelationsmatrix für die spätere Analyse (mit dem Unterbefehl `MATRIX`).

Vollständige Informationen zur Syntax finden Sie in der *SPSS Command Syntax Reference*.

Kapitel
30

Faktorenanalyse

Mit der Faktorenanalyse wird versucht, die zugrundeliegenden Variablen oder **Faktoren** zu bestimmen, welche die Korrelationsmuster innerhalb eines Satzes beobachteter Variablen erklären. Die Faktorenanalyse wird häufig zur Datenreduktion verwendet, indem wenige Faktoren identifiziert werden, welche den größten Teil der in einer großen Anzahl manifester Variablen aufgetretenen Varianz erklären. Die Faktorenanalyse kann auch zum Erzeugen von Hypothesen über kausale Mechanismen oder zum Sichten von Variablen für die anschließende Analyse verwendet werden (zum Beispiel, um vor einer linearen Regressionsanalyse Kollinearität zu erkennen).

Die Prozedur "Faktorenanalyse" bietet ein hohes Maß an Flexibilität:

- Es stehen sieben Methoden der Faktorextraktion zur Verfügung.
- Es sind fünf Rotationsmethoden verfügbar, einschließlich der direkten Oblimin-Methode und Promax-Methode für nicht orthogonale Rotationen.
- Für die Berechnung von Faktorwerten stehen drei Methoden zur Verfügung. Die Werte können für weitere Analysen als Variablen gespeichert werden.

Beispiel. Welche Einstellungen der befragten Personen liegen den gegebenen Antworten bei einer politischen Untersuchung zugrunde? Bei der Untersuchung der Korrelationen zwischen den Themen der Umfrage zeigen sich signifikante Überschneidungen zwischen verschiedenen Untergruppen von Themen. Fragen zu Steuern korrelieren gewöhnlich miteinander, ebenso wie Fragen zum Thema Bundeswehr und so weiter. Mit der Faktorenanalyse können Sie die Anzahl der zugrundeliegenden Faktoren untersuchen und in vielen Fällen die Bedeutung der Faktoren konzeptionell bestimmen. Zusätzlich können Sie für jeden Fall Faktorwerte berechnen lassen, die sich dann für weiterführende Analysen verwenden lassen. Zum Beispiel könnten Sie ein logistisches Regressionsmodell erstellen, um das Wahlverhalten auf der Grundlage von Faktorwerten vorherzusagen.

Statistiken. Für jede Variable: Anzahl gültiger Fälle, Mittelwert und Standardabweichung. Für jede Faktorenanalyse: Korrelationsmatrix der Variablen mit Signifikanzniveaus, Determinante, Inverse; reproduzierte Korrelationsmatrix mit Anti-Image; Anfangslösung (Kommunalitäten, Eigenwerte und Prozentsatz der erklärten Varianz); Kaiser-Meyer-Olkin-Maß für die Angemessenheit der Stichproben und Bartlett-Test auf Sphärizität; nicht rotierte Lösung mit Faktorladungen, Kommunalität und Eigenwerten; rotierte Lösung mit rotierter Mustermatrix und Transformationsmatrix. Für schiefe Rotationen: rotierte Muster- und Strukturmatrizen; Koeffizientenmatrix der Faktorwerte und Kovarianzmatrix des Faktors. Diagramme: Screeplot von Eigenwerten und Diagramm der Ladungen der ersten zwei oder drei Faktoren.

Daten. Die Variablen müssen auf dem **Intervall-** oder **Verhältnis-** Niveau quantitativ sein. Kategoriale Daten (wie beispielsweise Religion oder Geburtsland) sind für die Faktorenanalyse nicht geeignet. Daten, für welche die Korrelationskoeffizienten nach Pearson sinnvoll berechnet werden können, eignen sich gewöhnlich für eine Faktorenanalyse.

Annahmen. Die Daten sollten für jedes Variablenpaar in einer bivariaten Normalverteilung vorliegen. Beobachtungen müssen unabhängig sein. Im Modell der Faktorenanalyse ist festgelegt, daß Variablen durch gemeinsame Faktoren (die vom Modell geschätzten Faktoren) und eindeutige Faktoren (die sich nicht zwischen den beobachteten Variablen überschneiden) bestimmt sind. Die errechneten Schätzwerte basieren auf der Annahme, daß alle eindeutigen Faktoren weder miteinander noch mit den gemeinsamen Faktoren korrelieren.

Abbildung 30-1
Ausgabe der Faktorenanalyse

Deskriptive Statistiken

	Mittelwert	Standardabweichung	Analyse N
Durchschnittliche Lebenserwartung für Frauen	72.83	8.27	72
Kindersterblichkeit (Verstorbene pro 1000 Lebendgeburten)	35.132	32.222	72
Lesefähige Bevölkerung (%)	82.47	18.63	72
Geburtenrate pro 1000 Einwohner	24.375	10.552	72
Fruchtbarkeit: Durchschnittliche Kinderzahl	3.205	1.593	72
Stadtbevölkerung (%)	62.58	22.84	72
Log (Basis 10) von BIP_KOPF	3.5043	.6077	72
Bevölkerungswachstum (% pro Jahr)	1.697	1.156	72
Geburten/Verstorbene	3.5772	2.3131	72
Sterberate pro 1000 Einwohner	8.04	3.17	72
Log (Basise 10) der Einwohnerzahl	4.1526	.6865	72

Kommunalitäten

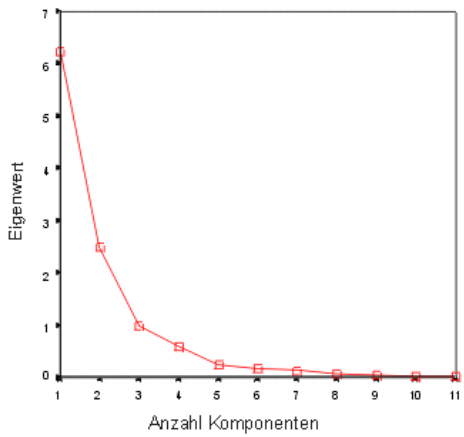
	Anfänglich	Extraktion
LEBERWF	1.000	.953
KINDSTER	1.000	.949
LESEN	1.000	.825
GEB_RT	1.000	.943
FRUCHTB	1.000	.875
URBAN	1.000	.604
LOG_BIP	1.000	.738
BEV_WACH	1.000	.945
G_ZU_V	1.000	.925
STERB_RT	1.000	.689
LOG_EINW	1.000	.292

Extraktionsmethode:
Hauptkomponentenanalyse.

Erklärte Gesamtvarianz

		Anfängliche Eigenwerte			Summen von quadrierten Faktorladungen für Extraktion			Rotierte Summe der quadrierten Ladungen		
		Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
Komponente	1	6.242	56.750	56.750	6.242	56.750	56.750	6.108	55.525	55.525
	2	2.495	22.685	79.435	2.495	22.685	79.435	2.630	23.910	79.435
	3	.988	8.986	88.421						
	4	.591	5.372	93.793						
	5	.236	2.142	95.935						
	6	.172	1.561	97.496						
	7	.124	1.126	98.622						
	8	6.958E-02	.633	99.254						
	9	4.460E-02	.405	99.660						
	10	2.443E-02	.222	99.882						
	11	1.302E-02	.118	100.000						

Extraktionsmethode: Hauptkomponentenanalyse.



Rotated Component Matrix

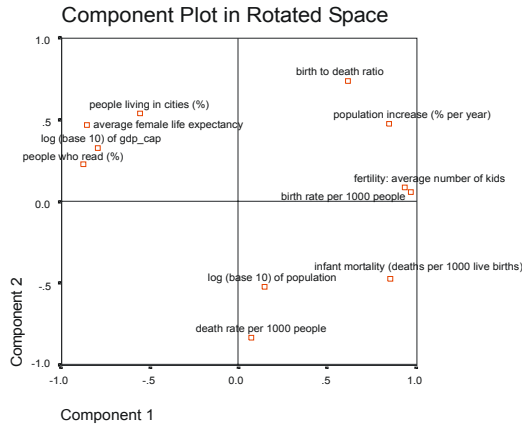
	Component	
	1	2
BIRTH_RT	.969	
FERTILTY	.931	
LITERACY	-.880	.226
LIFEEXPF	-.856	.469
BABYMORT	.853	-.469
POP_INCR	.847	.476
LOG_GDP	-.794	.327
URBAN	-.561	.539
DEATH_RT		-.827
B_TO_D	.614	.741
LOG_POP		-.520

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Komponententransformationsmatrix

		1	2
Komponente	1	.982	-.190
	2	.190	.982

Extraktionsmethode: Hauptkomponentenanalyse.
Rotationsmethode: Varimax mit Kaiser-Normalisierung.



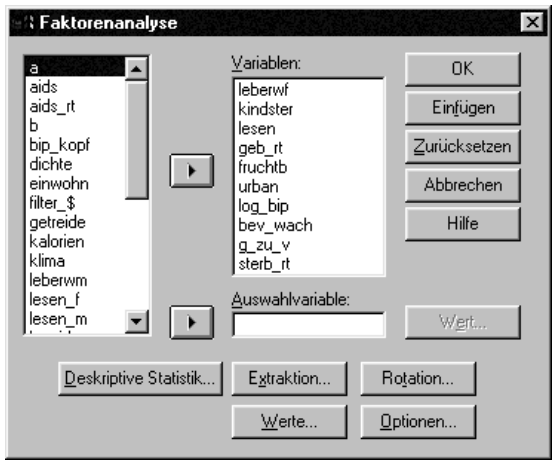
So lassen Sie eine Faktorenanalyse berechnen:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Dimensionsreduktion
  Faktorenanalyse…

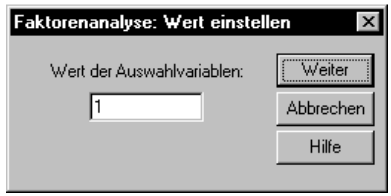
- Wählen Sie die Variablen für die Faktorenanalyse aus.

Abbildung 30-2
Dialogfeld "Faktorenanalyse"



Faktorenanalyse: Fälle auswählen

Abbildung 30-3
Dialogfeld "Faktorenanalyse: Wert einstellen"



So wählen Sie die Fälle für die Analyse aus:

- Wählen Sie eine Auswahlvariable aus.
- Klicken Sie auf Wert, um eine Ganzzahl als Auswahlvariable einzugeben.

Nur Fälle mit diesem Wert für die Auswahlvariable werden für die Faktorenanalyse verwendet.

Faktorenanalyse: Deskriptive Statistiken

Abbildung 30-4
Dialogfeld "Faktorenanalyse: Deskriptive Statistiken"

Statistik. Univariate Statistiken enthalten den Mittelwert, die Standardabweichung und die Anzahl gültiger Fälle für jede Variable. Die Anfangslösung zeigt die anfänglichen Kommunalitäten, Eigenwerte und den Prozentwert der erklärten Varianz an.

Korrelationsmatrix. Die verfügbaren Optionen sind Koeffizienten, Signifikanzniveaus, Determinante, Inverse, Reproduziert, Anti-Image sowie KMO und Bartlett-Test auf Sphärizität.

- **KMO und Bartlett-Test auf Sphärizität.** Das Kaiser-Meyer-Olkin-Maß für Angemessenheit der Stichproben überprüft, ob die partiellen Korrelationen zwischen Variablen klein sind. Der Bartlett-Test auf Sphärizität prüft, ob die Korrelationsmatrix eine Einheitsmatrix ist, wobei das Faktorenmodell in diesem Fall ungeeignet wäre.
- **Reproduziert.** Die geschätzte Korrelationsmatrix aus der Faktorlösung. Residuen (Differenz zwischen geschätzten und beobachteten Korrelationen) werden ebenfalls angezeigt.
- **Anti-Image.** Die Anti-Image-Korrelationsmatrix enthält die negativen Werte der partiellen Korrelationskoeffizienten. Die Anti-Image-Kovarianzmatrix enthält die negativen Werte der partiellen Kovarianzen. In einem guten Faktorenmodell sind die meisten außerhalb der Diagonalen liegenden Elemente klein. Das Maß der Stichprobeneignung einer Variablen wird auf der Diagonalen der Anti-Image-Korrelationsmatrix angezeigt.

Faktorenanalyse: Extraktion

Abbildung 30-5
Dialogfeld "Faktorenanalyse: Extraktion"

Methode. Hier kann die Methode der Faktorenextraktion festgelegt werden. Folgende Methoden sind verfügbar: Hauptkomponenten, ungewichtete kleinste Quadrate, verallgemeinerte kleinste Quadrate, Maximum Likelihood, Hauptachsen-Faktorenanalyse, Alpha-Faktorisierung und Image-Faktorisierung.

- **Hauptkomponentenanalyse (Factor Analysis).** Eine Methode zur Faktorextraktion. Sie wird verwendet, um unkorrelierte Linearkombinationen der beobachteten Variablen zu bilden. Die erste Komponente besitzt den größten Varianzanteil. Nachfolgende Komponenten erklären stufenweise kleinere Anteile der Varianz. Sie sind alle miteinander unkorreliert. Die Hauptkomponentenanalyse wird zur Ermittlung der Anfangslösung der Faktorenanalyse verwendet. Sie kann verwendet werden, wenn die Korrelationsmatrix singulär ist.
- **Ungewichtete kleinste Quadrate (Factor Analysis).** Eine Faktorextraktionsmethode, welche die Summe der quadrierten Differenzen zwischen der beobachteten und der reproduzierten Korrelationsmatrix unter Nichtberücksichtigung der Diagonalen minimiert.
- **Verallgemeinerte Methode der kleinsten Quadrate (Factor Analysis).** Eine Methode der Faktorextraktion, welche die Summe der quadrierten Abweichungen zwischen der beobachteten und der reproduzierten Korrelationsmatrix minimiert. Die Korrelationen werden mit dem inversen Wert der Eindeutigkeit gewichtet, so daß Variablen mit großer Eindeutigkeit schwach und solche mit kleiner Eindeutigkeit stärker gewichtet werden.
- **Maximum-Likelihood-Methode (Factor Analysis).** Eine Methode für die Faktorextraktion, die Parameterschätzer erzeugt, bei denen die Wahrscheinlichkeit am größten ist, daß sie die beobachtete Korrelationsmatrix erzeugt haben, wenn die Stichprobe aus einer multivariaten Normalverteilung stammt. Die Korrelationen werden durch die inverse Eindeutigkeit der Variablen gewichtet, und es wird ein iterativer Algorithmus eingesetzt.
- **Hauptachsen-Faktorenanalyse (Factor Analysis).** Eine Methode der Faktorextraktion aus der ursprünglichen Korrelationsmatrix, bei der die auf der Diagonalen befindlichen quadrierten Korrelationskoeffizienten als Anfangsschätzer der Kommunalitäten verwendet werden. Diese Faktorladungen werden benutzt, um neue Kommunalitäten zu schätzen, welche die alten Schätzer auf der Diagonalen ersetzen. Die Iterationen werden so lange fortgesetzt, bis die Änderungen in den Kommunalitäten von einer Iteration zur nächsten das Konvergenzkriterium der Extraktion erfüllen.
- **Alpha (Factor).** Eine Methode der Faktorextraktion, welche die Variablen in der Analyse als eine Stichprobe aus einer Grundgesamtheit aller potentiellen Variablen betrachtet. Dies vergrößert die Alpha-Reliabilität der Faktoren.
- **Image-Faktorisierung (Factor Analysis).** Eine Faktorextraktionsmethode, die von Guttman entwickelt wurde und auf der Imagetheorie basiert. Der gemeinsame Teil einer Variablen - partielles Image genannt - ist als ihre lineare Regression

auf die verbleibenden Variablen definiert, und nicht als eine Funktion von hypothetischen Faktoren.

Analysieren. Hier können Sie entweder eine Korrelationsmatrix oder eine Kovarianzmatrix festlegen.

- **Korrelationsmatrix.** Diese Funktion ist nützlich, wenn die Variablen in Ihrer Analyse anhand verschiedener Skalen gemessen werden.
- **Kovarianzmatrix.** Diese Funktion ist nützlich, wenn Sie die Faktorenanalyse auf mehrere Gruppen mit unterschiedlichen Varianzen für die einzelnen Variablen anwenden möchten.

Extrahieren. Sie können entweder alle Faktoren, deren Eigenwerte über einem festgelegten Wert liegen, oder eine festgelegte Anzahl von Faktoren beibehalten.

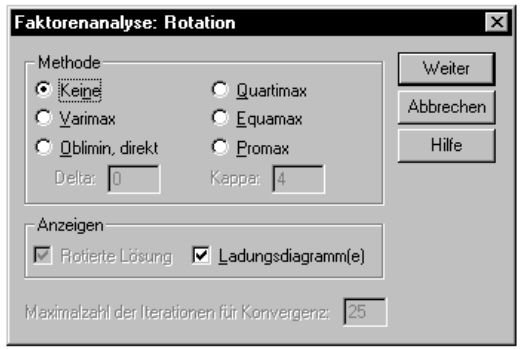
Anzeigen. Hier können Sie die nicht rotierte Faktorlösung und ein Screeplot der Eigenwerte anfordern.

- **Nicht rotierte Faktorlösung (Factor Analysis).** Zeigt unrotierte Faktorladungen (Faktormustermatrix), Kommunalitäten und Eigenwerte für die Faktorlösung an.
- **Screeplot.** Ein Diagramm der Varianz, die jedem Faktor zugeordnet ist. Es dient dazu, die Anzahl der Faktoren zu bestimmen, die behalten werden soll. Normalerweise zeigt das Diagramm einen deutlichen Bruch zwischen der starken Steigung der "großen" Faktoren und dem graduellen Verlauf der restlichen Faktoren (der "Geröllhalde").

Maximalzahl der Iterationen für Konvergenz. Hier können Sie für den Algorithmus eine Maximalzahl von Schritten zum Schätzen der Lösung festlegen.

Faktorenanalyse: Rotation

Abbildung 30-6
Dialogfeld "Faktorenanalyse: Rotation"



Methode. Hier können Sie die Methode der Faktor-Rotation auswählen. Die verfügbaren Methoden sind Varimax, Quartimax, Equamax, Promax oder Oblimin, direkt.

- **Varimax-Rotation (Factor Analysis).** Eine orthogonale Rotationsmethode, die die Anzahl der Variablen mit hohen Ladungen für jeden Faktor minimiert. Sie vereinfacht die Interpretation der Faktoren.
- **Methode "Oblimin, direkt".** Ein Verfahren zur schiefwinkligen (nichtorthogonalen) Rotation. Wenn Delta den Wert 0 annimmt (Standardeinstellung), sind die Ergebnisse am meisten schiefwinklig. Mit zunehmendem negativen Wert von Delta werden die Faktoren weniger schiefwinklig. Um den Standardwert von 0 zu überschreiben, geben Sie eine Zahl kleiner gleich 0,8 ein.
- **Quartimax-Rotation (Factor Analysis).** Eine Rotationsmethode, welche die Zahl der Faktoren minimiert, die zum Erklären aller Variablen benötigt werden. Sie vereinfacht die Interpretation der beobachteten Variablen.
- **Equamax-Rotation (Factor Analysis).** Eine Rotationsmethode, die eine Kombination zwischen der Varimax-Methode (sie vereinfacht Faktoren) und der Quartimax-Methode (sie vereinfacht Variablen) darstellt. Die Anzahl der Variablen mit hohen Ladungen auf einen Faktor sowie die Anzahl der Faktoren, die benötigt werden, um eine Variable zu erklären, werden minimiert.
- **Promax-Rotation (Factor Analysis).** Eine schiefe Rotation, bei der Faktoren korreliert sein dürfen. Diese Rotaion kann schneller berechnet werden als eine direkte Oblimin-Rotation und ist daher nützlich für große Datendateien.

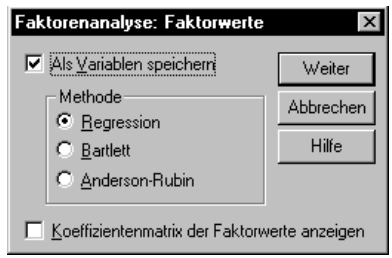
Anzeigen. Hiermit können Sie eine Ausgabe für die rotierte Lösung sowie Ladungsdiagramme für die ersten zwei oder drei Faktoren einbeziehen.

- **Rotierte Lösung (Factor Analysis).** Um eine rotierte Lösung zu erhalten, muß eine Rotationsmethode ausgewählt sein. Für orthogonale Rotationen werden die rotierte Mustermatrix und Faktortransformationsmatrix angezeigt. Für schiefe Rotationen werden Muster, Struktur und Faktorkorrelationsmatrix angezeigt.
- **Diagramm der Faktorladungen.** Dreidimensionales Diagramm der Faktorladungen für die ersten drei Faktoren. Für eine Lösung mit zwei Faktoren wird ein dreidimensionales Diagramm angezeigt. Das Diagramm wird nicht angezeigt, wenn nur ein Faktor extrahiert wird. Auf Wunsch zeigen die Diagramme rotierte Lösungen an.

Maximalzahl der Iterationen für Konvergenz. Hier können Sie eine Maximalzahl von Schritten zum Durchführen der Rotation für den Algorithmus festlegen.

Faktorenanalyse: Faktorwerte

Abbildung 30-7
Dialogfeld "Faktorenanalyse: Faktorwerte"



Als Variablen speichern. Hiermit wird für jeden Faktor in der endgültigen Lösung eine neue Variable erstellt. Wählen Sie eine der folgenden Methoden für die Berechnung der Faktorwerte aus: Regression, Bartlett oder Anderson-Rubin.

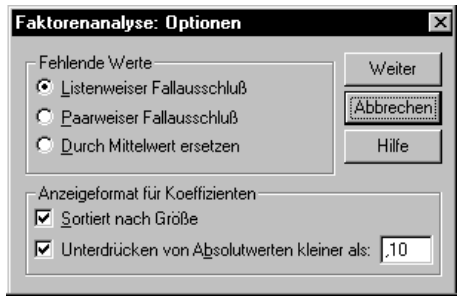
- **Regressionsmethode (Factor Analysis).** Eine Methode, um Koeffizienten für Faktorwerte zu schätzen. Die Faktorwerte haben einen Mittelwert von 0 und eine Varianz, die der quadrierten Mehrfachkorrelation zwischen den geschätzten und den wahren Faktorwerten entspricht. Die Werte können korreliert sein, selbst wenn die Faktoren orthogonal sind.

- **Barlett-Werte.** Eine Methode zum Schätzen von Koeffizienten für Faktorwerte. Die erzeugten Faktorwerte haben einen Mittelwert von 0. Die Quadratsumme der eindeutigen Faktoren über dem Variablenbereich wird minimiert.
- **Anderson-Rubin-Methode (Factor Analysis).** Eine Methode zur Berechnung von Faktorwerten; eine Modifizierung der Bartlett-Methode, die die Orthogonalität der geschätzten Faktoren gewährleistet. Die berechneten Werte haben einen Mittelwert von 0 und eine Standardabweichung von 1 und sind unkorreliert.

Koeffizientenmatrix der Faktorwerte anzeigen. Hiermit werden die Koeffizienten angezeigt, mit denen die Variablen multipliziert werden, um Faktorwerte zu erhalten. Hiermit werden auch die Korrelationen zwischen Faktorwerten angezeigt.

Faktorenanalyse: Optionen

Abbildung 30-8
Dialogfeld "Faktorenanalyse: Optionen"



Fehlende Werte. Hier können Sie festlegen, wie fehlende Werte behandelt werden. Es ist möglich, die Fälle **listenweise** auszuschließen, **paarweise** auszuschließen oder durch den Mittelwert zu ersetzen.

Anzeigeformat für Koeffizienten. Hiermit können Sie Einstellungen für Aspekte der Ausgabematrix vornehmen. Sie können die Koeffizienten nach Größe sortieren lassen und Koeffizienten mit absoluten Werten unterdrücken, die kleiner als der festgelegte Wert sind.

Zusätzliche Funktionen beim Befehl FACTOR

Mit der SPSS-Befehlssprache verfügen Sie über die folgenden zusätzlichen Möglichkeiten:

- Angeben von Konvergenzkriterien für die Iteration während der Extraktion und Rotation.
- Angeben von einzelnen rotierten Faktordiagrammen.
- Angeben der Anzahl der zu speichernden Faktorwerte.
- Angeben der Diagonalwerte für die Hauptachsen-Faktorenanalyse.
- Schreiben der Korrelationsmatrizen oder der Faktorladungs-Matrizen auf die Festplatte für eine spätere Analyse.
- Einlesen und Analysieren von Korrelationsmatrizen oder Faktorladungs-Matrizen.

Vollständige Informationen zur Syntax finden Sie in der *SPSS Command Syntax Reference*.

Kapitel
31
Auswählen einer Prozedur zum Durchführen einer Clusteranalyse

Clusteranalysen können mit den Prozeduren "Two-Step-Clusteranalyse", "Hierarchische Clusteranalyse" oder "Clusterzentrenanalyse" durchgeführt werden. In jeder Prozedur wird ein anderer Algorithmus zum Erstellen von Clustern eingesetzt, und jede Prozedur verfügt über Optionen, die in den jeweils anderen Prozeduren nicht verfügbar sind.

Two-Step-Clusteranalyse. In vielen Fällen ist die Prozedur "Two-Step-Clusteranalyse" die beste Wahl. Sie bietet die folgenden speziellen Funktionen:

- Automatische Auswahl der optimalen Anzahl von Clustern sowie Maße, die bei der Auswahl des Cluster-Modells helfen
- Gleichzeitiges Erstellen von Cluster-Modellen mit kategorialen und stetigen Variablen
- Speichern des Cluster-Modells in einer externen XML-Datei und anschließendem Einlesen dieser Datei und Aktualisieren des Cluster-Modells mit neuen Daten.

Außerdem können von der Prozedur "Two-Step-Clusteranalyse" auch umfangreiche Datendateien analysiert werden.

Hierarchische Clusteranalyse. Die Prozedur "Hierarchische Clusteranalyse" ist auf kleinere Datendateien begrenzt (mehrere Hundert zu gruppierende Objekte), bietet jedoch die folgenden speziellen Funktionen:

- Möglichkeit der Zusammenfassung von Fällen oder Variablen in Clustern
- Funktion zum Berechnen eines Bereichs möglicher Lösungen und zum Speichern der Cluster-Zugehörigkeiten für jede dieser Lösungen
- Verschiedene Methoden zur Clusterbildung, Transformation von Variablen und Messung der Unähnlichkeit zwischen Clustern

Mit der Prozedur "Hierarchische Clusteranalyse" können Intervallvariablen (stetige Variablen), Zählvariablen oder binäre Variablen analysiert werden, wobei alle für die Prozedur ausgewählten Variablen jeweils denselben Typ aufweisen müssen.

Clusterzentrenanalyse. Die Prozedur "Clusterzentrenanalyse" ist auf stetige Daten beschränkt und setzt eine Festlegung der Cluster-Anzahl voraus, bietet jedoch die folgenden speziellen Funktionen:

- Funktion zum Speichern der Distanz vom Clusterzentrum für jedes Objekt
- Funktion zum Einlesen der anfänglichen Clusterzentren aus einer externen SPSS-Datei und zum Speichern der endgültigen Clusterzentren in dieser Datei

Außerdem können von der Prozedur "Clusterzentrenanalyse" auch umfangreiche Datendateien analysiert werden.

Kapitel
32

Two-Step-Clusteranalyse

Bei der Two-Step-Clusteranalyse handelt es sich um eine explorative Prozedur zum Ermitteln von natürlichen Gruppierungen (Clustern) innerhalb eines Datensatzes, die andernfalls nicht erkennbar wären. Der von der Prozedur verwendete Algorithmus verfügt über vielfältige nützliche Funktionen, durch die er sich von traditionellen Cluster-Methoden unterscheidet:

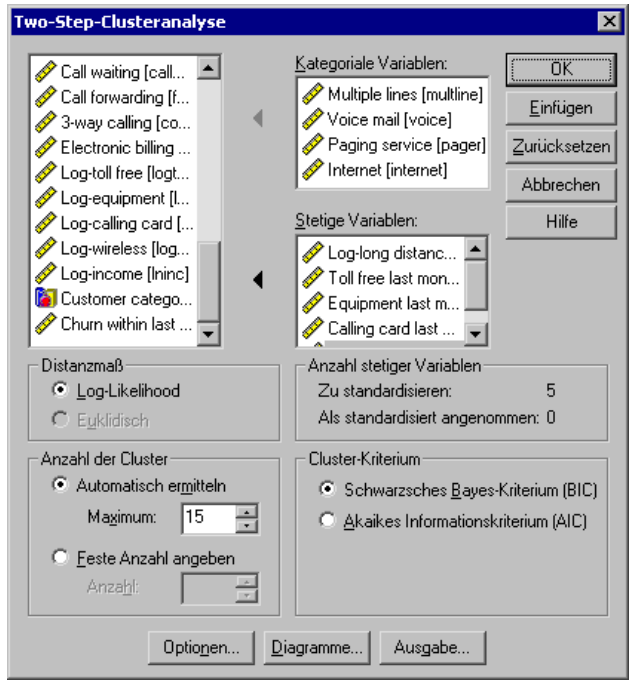
- **Verarbeitung von kategorialen und stetigen Variablen.** Die Annahme der Unabhängigkeit der Variablen ermöglicht eine kombinierte multinomiale Normalverteilung für kategoriale und stetige Variablen.
- **Automatische Auswahl der Cluster-Anzahl.** Durch den Vergleich der Werte eines Modellauswahlkriteriums in verschiedenen Clusteranalysen kann die optimale Anzahl der Cluster von der Prozedur automatisch bestimmt werden.
- **Skalierbarkeit.** Durch das Zusammenfassen der Datensätze in einem Clusterfunktionsbaum (CF-Baum) können mit dem Two-Step-Algorithmus sehr große Datendateien analysiert werden.

Beispiel. In Einzel- und Fachhandel werden Cluster-Methoden regelmäßig auf Daten angewendet, die Kaufgewohnheiten, Geschlecht, Alter und Einkommensniveau der Kundschaft beschreiben. Ziel der Analyse ist eine Ausrichtung der unternehmenseigenen Marketing- und Produktentwicklungsstrategien auf einzelne Konsumentengruppen, um Umsatzsteigerungen und Markentreue zu erreichen.

Statistiken. Mit dieser Prozedur werden Informationskriterien (AIC oder BIC) nach Anzahl der Cluster sowie Cluster-Häufigkeiten und deskriptive Statistiken nach Cluster für die abschließende Clusteranalyse erstellt.

Diagramme. Mit dieser Prozedur werden Balken- und Kreisdiagramme für Cluster-Häufigkeiten sowie Wichtigkeitsdiagramme für Variablen erstellt.

Abbildung 32-1
Dialogfeld "Two-Step-Clusteranalyse"



Distanzmaß. Mit dieser Auswahl legen Sie fest, wie Ähnlichkeiten zwischen zwei Clustern verarbeitet werden.

- **Log-Likelihood.** Mit dem Likelihood-Maß wird eine Wahrscheinlichkeitsverteilung für die Variablen vorgenommen. Bei stetigen Variablen wird von einer Normalverteilung, bei kategorialen Variablen von einer multinomialen Verteilung ausgegangen. Bei allen Variablen wird davon ausgegangen, daß sie unabhängig sind.
- **Euklidisch.** Das Euklidische Maß bezeichnet die "gerade" Distanz zwischen zwei Clustern. Es kann nur dann verwendet werden, wenn es sich bei sämtlichen Variablen um stetige Variablen handelt.

Anzahl der Cluster. Mit dieser Auswahl können Sie angeben, wie die Anzahl der Cluster bestimmt werden soll.

- **Automatisch ermitteln.** Mit dieser Prozedur wird das im Gruppenfeld "Cluster-Kriterium" angegebene Kriterium verwendet, um automatisch die "beste" Anzahl der Cluster zu ermitteln. Sie haben die Möglichkeit, eine positive Ganzzahl für die Höchstzahl der Cluster anzugeben, die von der Prozedur berücksichtigt werden sollen.
- **Feste Anzahl angeben.** Ermöglicht das Festlegen der Anzahl der Cluster für die Analyse. Geben Sie eine positive Ganzzahl ein.

Anzahl stetiger Variablen. Dieses Gruppenfeld enthält eine Zusammenfassung der Standardeinstellungen, die im Dialogfeld "Optionen" für stetige Variablen vorgenommen wurden. Für weitere Informationen siehe "Two-Step-Clusteranalyse: Optionen" auf S. 537.

Cluster-Kriterium. Mit dieser Auswahl legen Sie fest, wie die Anzahl der Cluster vom automatischen Cluster-Algorithmus bestimmt wird. Angegeben werden kann entweder das Bayes-Informationskriterium (BIC) oder das Akaikes-Informationskriterium (AIC).

Daten. Mit dieser Prozedur können sowohl stetige als auch kategoriale Variablen analysiert werden. Die Fälle bilden dabei die Objekte, die gruppiert werden sollen, während die Variablen die Attribute darstellen, auf deren Grundlage die Gruppierung erfolgt.

Fallreihenfolge. Beachten Sie, daß der Cluster-Funktionsbaum und die endgültige Lösung ggf. von der Reihenfolge der Fälle abhängig sein können. Um die Auswirkungen der Reihenfolge zu minimieren, mischen Sie die Fälle in zufälliger Reihenfolge. Prüfen Sie daher die Stabilität einer bestimmten Lösung, indem Sie verschiedene Lösungen abrufen, bei denen die Fälle in einer unterschiedlichen, zufällig ausgewählten Reihenfolge sortiert sind. In schwierigen Situationen mit äußerst umfangreichen Dateien führen Sie statt dessen mehrere Läufe aus, bei denen eine Stichprobe der Fälle in unterschiedlicher, zufälliger Reihenfolge angeordnet ist.

Annahmen. Das Likelihood-Distanzmaß geht davon aus, daß die Variablen im Clustermodell unabhängig sind. Außerdem wird für stetige Variablen eine Normal- bzw. Gauß-Verteilung und für kategoriale Variable eine multinomiale Verteilung vorausgesetzt. Empirische interne Tests zeigen, daß die Prozedur wenig anfällig gegenüber Verletzungen hinsichtlich der Unabhängigkeitsannahme und der Verteilungsannahme ist. Dennoch sollten Sie darauf achten, wie genau diese Voraussetzungen erfüllt sind.

Mit der Prozedur Bivariate Korrelationen können Sie die Unabhängigkeit zwischen zwei stetigen Variablen überprüfen. Mit der Prozedur Kreuztabellen können Sie die Unabhängigkeit zwischen zwei kategorialen Variablen überprüfen. Mit der Prozedur Mittelwerte können Sie die Unabhängigkeit zwischen einer stetigen und einer kategorialen Variablen überprüfen. Mit der Prozedur Explorative Datenanalyse prüfen Sie die Normalverteilung einer stetigen Variablen. Mit der Prozedur Chi-Quadrat-Test überprüfen Sie, ob eine kategoriale Variable über eine bestimmte multinomiale Verteilung verfügt.

So lassen Sie eine Two-Step-Clusteranalyse berechnen

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Klassifizieren
    Two-Step-Clusteranalyse...

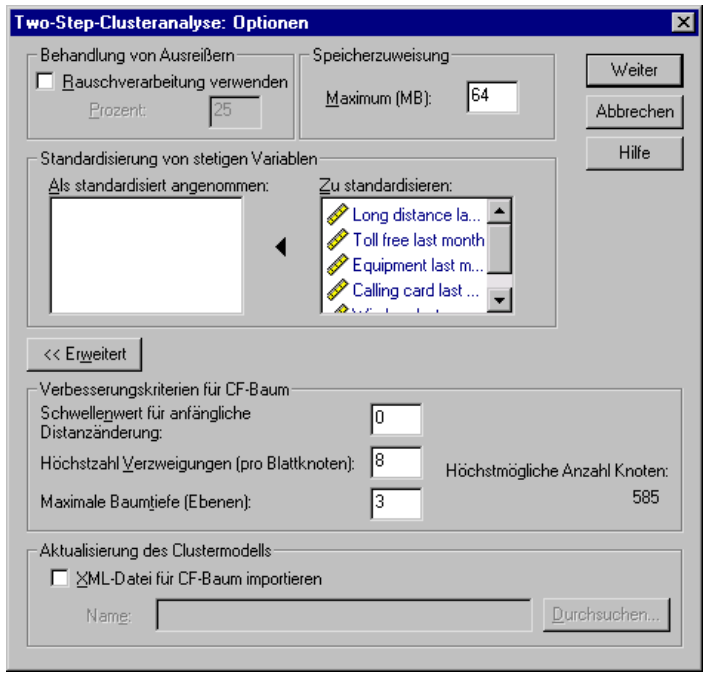
▶ Wählen Sie mindestens eine kategoriale oder stetige Variable aus.

Die folgenden Optionen sind verfügbar:

- Anpassen der Kriterien für die Erstellung der Cluster
- Auswählen der Einstellungen für die Rauschverarbeitung, Speicherzuweisung, Variablenstandardisierung und Eingabe des Clustermodells
- Anfordern von optionalen Tabellen und Diagrammen
- Speichern der Modellergebnisse in der Arbeitsdatei oder in einer externen XML-Datei

Two-Step-Clusteranalyse: Optionen

Abbildung 32-2
Dialogfeld "Two-Step-Clusteranalyse: Optionen"



Behandlung von Ausreißern. Mit diesem Gruppenfeld können Sie Ausreißer während des Füllvorgangs des CF-Baums bei der Clusteranalyse gesondert behandeln. Der CF-Baum ist vollständig, wenn keine weiteren Fälle in einem Blattknoten aufgenommen werden können und kein Blattknoten mehr aufgeteilt werden kann.

- Wenn während des Füllvorgangs des CF-Baums eine Rauschverarbeitung stattfinden soll, wird der CF-Baum neu gebildet, nachdem Fälle von wenig besetzten Blättern auf einem "Rauschblatt" positioniert worden sind. Ein Blatt wird als wenig besetzt betrachtet, wenn es weniger Fälle als den angegebenen Prozentsatz der maximalen Blattgröße enthält. Nach der Neubildung des Baums

können gegebenenfalls noch Ausreißer im CF-Baum positioniert werden. Andernfalls werden die Ausreißer verworfen.

- Wenn während des Füllvorgangs des CF-Baums keine Rauschverarbeitung stattfinden soll, wird der Baum unter Verwendung eines größeren Schwellenwerts für die Distanzänderung neu gebildet. Nach der abschließenden Clusteranalyse werden die Werte, die keinem Cluster zugewiesen werden konnten, als Ausreißer bezeichnet. Der Ausreißer-Cluster erhält die Identifikationsnummer –1 und wird nicht in die Auszählung der Anzahl von Clustern aufgenommen.

Speicherzuweisung. In diesem Gruppenfeld können Sie den maximalen Speicherplatz in MB angeben, der vom Cluster-Algorithmus verwenden soll. Wenn der für die Prozedur erforderliche Speicherplatz den maximalen Speicherplatz übersteigt, wird die Festplatte zum Speichern der Daten verwendet, die nicht in den Arbeitsspeicher passen. Geben Sie eine Zahl größer oder gleich 4 ein.

- Den größtmöglichen Wert, den Sie für Ihr System angeben können, erfahren Sie bei Ihrem Systemadministrator.
- Wenn dieser Wert zu niedrig ist, kann die Anzahl der Cluster unter Umständen nicht ordnungsgemäß ermittelt werden.

Variablenstandardisierung. Mit dem Cluster-Algorithmus werden standardisierte stetigen Variablen analysiert. Alle stetigen Variablen, die nicht standardisiert sind, sollten in der Liste “Zu standardisieren” verbleiben. Um Zeit und Verarbeitungsaufwand zu sparen, können Sie alle bereits standardisierten stetigen Variablen in der Liste “Als standardisiert angenommen” auswählen.

Erweiterte Optionen

Verbesserungskriterien für CF-Baum. Die folgenden Einstellungen für den Cluster-Algorithmus gelten insbesondere für den CF-Baum und sollten nur nach sorgfältiger Prüfung geändert werden:

- **Schwellenwert für anfängliche Distanzänderung.** Hierbei handelt es sich um den anfänglichen Schwellenwert, der zum Erstellen des CF-Baums verwendet wird. Wenn das Hinzufügen eines gegebenen Falls zu einem Blatt des CF-Baums eine Dichte unterhalb dieses Schwellenwerts ergibt, wird das Blatt nicht geteilt. Wenn die Dichte den Schwellenwert überschreitet, wird das Blatt geteilt.

- **Höchstzahl Verzweigungen (pro Blattknoten).** Hierbei handelt es sich um die maximale Anzahl an untergeordneten Knoten, über die ein Blattknoten verfügen kann.
- **Maximale Baumtiefe.** Die maximale Anzahl an Ebenen, über die ein CF-Baum verfügen kann.
- **Höchstmögliche Anzahl Knoten.** Gibt die maximale Anzahl an CF-Baumknoten an, die von der Prozedur anhand der Gleichung $(b^{d+1} - 1) / (b - 1)$ potentiell erstellt werden können, wobei b für die Höchstzahl der Verzweigungen und d für die maximale Baumtiefe steht. Beachten Sie, daß ein extrem großer CF-Baum die Systemressourcen stark belastet und somit die Prozedurleistung beinträchtigen kann. Die Mindestanforderung pro Knoten beträgt 16 Bytes.

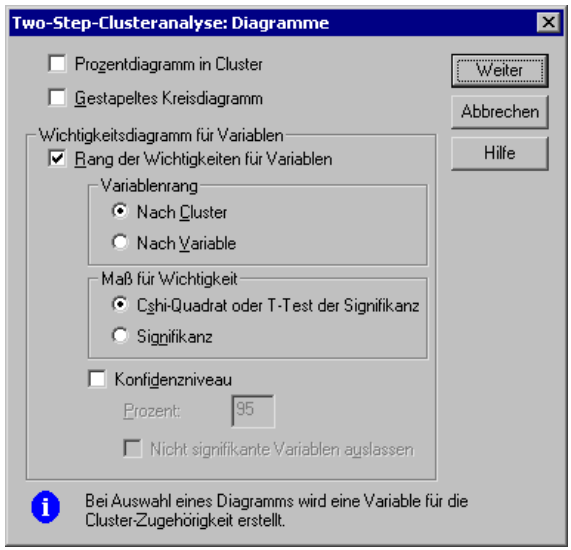
Aktualisierung des Clustermodells. Mit diesem Gruppenfeld können Sie ein Clustermodell importieren und aktualisieren, das in einer vorangegangenen Analyse erstellt wurde. Die Eingabedatei enthält den CF-Baum im XML-Format. Das Modell wird dann mit den Daten der aktiven Datei aktualisiert. Die Variablennamen müssen im Hauptdialogfeld in der Reihenfolge ausgewählt werden, in der sie in der vorangegangenen Analyse angegeben wurden. Die XML-Datei bleibt unverändert, es sei denn, Sie speichern die neuen Modelldaten unter demselben Dateinamen. Für weitere Informationen siehe “Two-Step-Clusteranalyse: Ausgabe” auf S. 541.

Bei einer Aktualisierung des Clustermodells werden zur Erstellung des CF-Baums dieselben Optionen verwendet, die für das ursprüngliche Modell gelten. Genauer gesagt werden die Optionen für Distanzmaß, Rauschverarbeitung, Speicherzuweisung und Verbesserungskriterien für den CF-Baum aus dem gespeicherten Modell übernommen, wobei die in den Dialogfeldern für diese Optionen vorgenommenen Einstellungen ignoriert werden.

Anmerkung: Beim Ausführen einer Aktualisierung des Clustermodells wird von der Prozedur vorausgesetzt, dass keiner der ausgewählten Fälle in der Arbeitsdatei für die Erstellung des ursprünglichen Clustermodells verwendet wurde. Außerdem gilt die Annahme, daß die Fälle für die Modellaktualisierung der gleichen Grundgesamtheit entstammen wie die Fälle, die zur Erstellung des ursprünglichen Modells verwendet wurden. Das heißt, es wird angenommen, daß die Mittelwerte und Varianzen der stetigen Variablen sowie die Ebenen der kategorialen Variablen in beiden Fallgruppen identisch sind. Wenn Ihre “neuen” und “alten” Fallgruppen aus heterogenen Grundgesamtheiten stammen, müssen Sie die Two-Step-Clusteranalyse für eine Kombination der beiden Fallgruppen ausführen, um optimale Ergebnisse zu erzielen.

Two-Step-Clusteranalyse: Diagramme

Abbildung 32-3
Dialogfeld "Two-Step-Clusteranalyse: Diagramme"



Prozentdiagramm in Cluster. Hierbei handelt es sich um Diagramme, in denen die Variation der einzelnen Variablen innerhalb eines Clusters angezeigt wird. Für jede kategoriale Variable wird ein gruppiertes Balkendiagramm erstellt, in dem Kategorienhäufigkeiten nach Cluster-ID angezeigt werden. Für jede stetige Variable wird ein Fehlerbalkendiagramm erstellt, in dem Fehlerbalken nach Cluster-ID angezeigt werden.

Gestapeltes Kreisdiagramm. Zeigt ein Kreisdiagramm an, das den Prozentsatz und die Häufigkeit der Beobachtungen innerhalb der einzelnen Cluster darstellt.

Wertigkeitsdiagramme für Variablen. Zeigt zahlreiche unterschiedliche Diagramme an, die die Wichtigkeit der einzelnen Variable in den einzelnen Clustern darstellen. In der Ausgabe werden die einzelnen Variablen nach Wichtigkeitsrang sortiert.

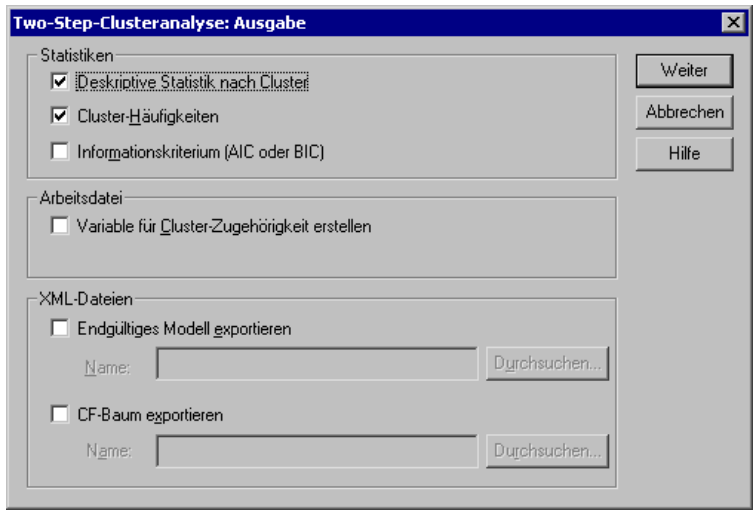
- **Variablenrang.** Mit dieser Option wird festgelegt, für jedes Cluster (Nach Cluster) oder für jede Variable (Nach Variable) ein Diagramm erstellt werden soll.
- **Maß für Wichtigkeit.** Mit dieser Option können Sie festlegen, welches Wichtigkeitsmaß für die Variablen grafisch dargestellt werden soll. Chi-Quadrat oder T-Test der Signifikanz gibt eine Pearson-Chi-Quadrat-Statistik als die

Wichtigkeit einer kategorialen Variable und eine *T*-Statistik als Wichtigkeit einer stetigen Variable aus. Signifikanz gibt eins minus *P*-Wert als Test auf Gleichheit der Mittelwerte für eine stetige Variable und die erwartete Häufigkeit im gesamten Daten-Set für eine kategoriale Variable aus.

- **Konfidenzniveau.** Mit dieser Option können Sie das Konfidenzniveau des Tests auf Gleichheit für die Verteilung einer Variablen innerhalb eines Clusters im Vergleich zur Gesamtverteilung der Variablen festlegen. Geben Sie eine Zahl ein, die kleiner als 100 und größer oder gleich 50 ist. Der Wert des Konfidenzniveaus wird als vertikale Linie in den Wichtigkeitsdiagrammen für Variablen dargestellt, wenn die Diagramme für Variablen erstellt werden oder wenn das Signifikanzmaß grafisch dargestellt wird.
- **Nicht signifikante Variablen auslassen.** Variablen, die für das angegebene Konfidenzniveau nicht signifikant sind, werden in den Wichtigkeitsdiagrammen für Variablen nicht angezeigt.

Two-Step-Clusteranalyse: Ausgabe

Abbildung 32-4
Dialogfeld "Two-Step-Clusteranalyse: Ausgabe"



Statistik. In diesem Gruppenfeld können Sie Optionen für die Anzeige von Tabellen mit den Ergebnissen der Clusterananlyse einstellen. Tabellen mit deskriptiven Statistiken und Cluster-Häufigkeiten eignen sich zur Darstellung des endgültigen

Clustermodells, während in der Tabelle mit Informationskriterien Ergebnisse für eine Reihe verschiedener Cluster-Lösungen angezeigt werden.

- **Deskriptive Statistik nach Cluster.** Zeigt zwei Tabellen an, die die Variablen in den einzelnen Clustern beschreiben. In der einen Tabelle werden die Mittelwerte und Standardabweichungen der stetigen Variablen nach Cluster erfaßt. In der anderen Tabelle werden die Häufigkeiten der kategorialen Variablen nach Cluster erfaßt.
- **Cluster-Häufigkeiten.** Zeigt eine Tabelle an, in der die Anzahl der Beobachtungen in den einzelnen Clustern erfaßt wird.
- **Informationskriterium (AIC oder BIC).** Zeigt eine Tabelle mit den Werten von AIC oder BIC für eine unterschiedliche Anzahl von Clustern an, je nachdem, welches Kriterium im Hauptdialogfeld ausgewählt wurde. Diese Tabelle wird lediglich dann bereitgestellt, wenn die Anzahl der Cluster automatisch festgelegt wurde. Bei einer festen Anzahl von Clustern wird die Einstellung ignoriert und die Tabelle nicht bereitgestellt.

Arbeitsdatei. Mit diesem Gruppenfeld können Sie Variablen in der Arbeitsdatei speichern.

- **Variable für Cluster-Zugehörigkeit erstellen.** Diese Variable enthält für jeden Fall eine Cluster-Identifikationsnummer. Der Name dieser Variablen lautet *tsc_n*, wobei *n* eine positive Ganzzahl ist, die auf die Ordinalzahl der Arbeitsdatei hinweist, die von dieser Prozedur in einer gegebenen Sitzung gespeichert wurde.

XML-Dateien. Das endgültige Clustermodell und der CF-Baum sind zwei Arten von Ausgabedateien, die als XML-Format exportiert werden können.

- **Endgültiges Modell exportieren.** Das endgültige Clustermodell wird in die angegebene Datei exportiert. SmartScore und SPSS Server (gesondertes Produkt) können anhand dieser Modelldatei die Modellinformationen zu Bewertungszwecken auf andere Datendateien anwenden.
- **CF-Baum exportieren.** Mit dieser Option können Sie den aktuellen Stand des Cluster-Baums speichern und zu einem späteren Zeitpunkt mit neuen Daten aktualisieren.

Kapitel
33

Hierarchische Clusteranalyse

Mit diesem Verfahren wird anhand ausgewählter Merkmale versucht, relativ homogene Fallgruppen oder Variablen zu identifizieren. Dabei wird ein Algorithmus eingesetzt, der für jeden Fall, oder für jede Variable, einen separaten Cluster bildet und die Cluster so lange kombiniert, bis nur noch einer zurückbleibt. Sie können einfache Variablen analysieren oder aus einer Vielfalt von Transformationen auswählen, welche die Daten standardisieren. Distanz- oder Ähnlichkeitsmaße werden durch die Prozedur "Ähnlichkeiten" erzeugt. Für jeden Schritt werden Statistiken angezeigt, um Sie bei der Auswahl der besten Lösung zu unterstützen.

Beispiel. Können Gruppen von verschiedenen Fernseh-Shows identifiziert werden, die ein ähnliches Publikum ansprechen? Mit Hilfe der hierarchischen Clusteranalyse können Sie die Fernseh-Shows (Fälle) anhand der Merkmale der Zuschauer in homogene Gruppen (Cluster) aufteilen. Damit lassen sich beispielsweise Marktsegmente identifizieren. Sie können außerdem Städte (Fälle) in homogene Gruppen clustern, so daß vergleichbare Städte zum Testen verschiedener Marketingstrategien ausgewählt werden können.

Statistiken. Zuordnungsübersicht, Distanz- oder Ähnlichkeitsmatrix und Cluster-Zugehörigkeit für eine einzelne Lösung oder einen Bereich von Lösungen. Diagramme: Dendrogramme und Eiszapfendiagramme.

Daten. Bei den Variablen kann es sich um quantitative Daten, binäre Daten oder Häufigkeitsdaten handeln. Die Skalierung der Variablen spielt eine wichtige Rolle. Unterschiede in der Skalierung können sich auf Ihre Cluster-Lösung(en) auswirken. Wenn Ihre Variablen sehr unterschiedlich skaliert sind, eine also beispielsweise in Dollar und die andere in Jahren angegeben wird, empfiehlt sich die Standardisierung. (Die Prozedur "Hierarchische Clusteranalyse" kann dies automatisch durchführen.)

Fallreihenfolge. Wenn gebundene Distanzen oder Ähnlichkeiten in den Eingabedaten vorliegen (oder beim Verbinden in den aktualisierten Clustern auftreten), ist die resultierende Cluster-Lösung ggf. abhängig von der Reihenfolge der Fälle in

der Datei. Prüfen Sie daher die Stabilität einer bestimmten Lösung, indem Sie verschiedene Lösungen abrufen, bei denen die Fälle in einer unterschiedlichen, zufällig ausgewählten Reihenfolge sortiert sind.

Annahmen. Die verwendeten Distanz- und Ähnlichkeitsmaße müssen für die analysierten Daten geeignet sein. Weitere Informationen zur Auswahl der Distanz- und Ähnlichkeitsmaße finden Sie unter der Prozedur "Ähnlichkeiten". Außerdem sollten Sie alle relevanten Variablen in Ihre Analyse einschließen. Das Weglassen einflußreicher Variablen kann zu irreführenden Lösungen führen. Da es sich bei der hierarchischen Clusteranalyse um eine explorative Methode handelt, sollten die Ergebnisse als vorläufig gelten, bis diese durch eine unabhängige Stichprobe bestätigt werden.

Abbildung 33-1
Ausgabe der "Hierarchischen Clusteranalyse"

Agglomeration Schedule

		Cluster Combined			Stage Cluster First Appears		
		Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Next Stage
Stage	1	11	12	.112	0	0	2
	2	6	11	.132	0	1	4
	3	7	9	.185	0	0	5
	4	6	8	.227	2	0	7
	5	7	10	.274	3	0	7
	6	1	3	.423	0	0	10
	7	6	7	.438	4	5	14
	8	13	14	.484	0	0	15
	9	2	5	.547	0	0	11
	10	1	4	.691	6	0	11
	11	1	2	1.023	10	9	13
	12	15	16	1.370	0	0	13
	13	1	15	1.716	11	12	14
	14	1	6	2.642	13	7	15
	15	1	13	4.772	14	8	0

Cluster Membership

		Label	4 Clusters	3 Clusters	2 Clusters
Case	1	Argentina	1	1	1
	2	Brazil	1	1	1
	3	Chile	1	1	1
	4	Domincan R.	1	1	1
	5	Indonesia	1	1	1
	6	Austria	2	2	1
	7	Canada	2	2	1
	8	Denmark	2	2	1
	9	Italy	2	2	1
	10	Japan	2	2	1
	11	Norway	2	2	1
	12	Switzerland	2	2	1
	13	Bangladesh	3	3	2
	14	India	3	3	2
	15	Bolivia	4	1	1
	16	Paraguay	4	1	1

Vertical Icicle

Number of clusters	14:India	14	13:Banglade	13	10:Japan	10	9:Italy	9	7:Canada	7	8:Denmark	8	12:Switzerlan	12	11:Norway	11	6:Austria	6	16:Paraguay	16	15:Bolivia	15	5:Indonesia	5	2:Brazil	2	4:Dominican	4	3:Chile	3	1:Argentina
	Case																														
1	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
3	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X
4	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X		X	X	X	X	X	X	X	X
5	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X		X	X	X	X	X	X	X	X	X
6	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X		X	X	X		X	X	X	X	X
7	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X		X	X	X		X		X	X	X
8	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X		X		X		X		X	X	X
9	X		X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X		X		X		X		X	X	X
10	X		X		X	X	X	X	X		X	X	X	X	X	X	X	X	X		X		X		X		X		X	X	X
11	X		X		X	X	X	X	X		X	X	X	X	X	X	X	X	X		X		X		X		X		X		X
12	X		X		X		X	X	X		X	X	X	X	X	X	X	X	X		X		X		X		X		X		X
13	X		X		X		X	X	X		X		X	X	X	X	X	X	X		X		X		X		X		X		X
14	X		X		X		X		X		X		X	X	X	X	X	X	X		X		X		X		X		X		X
15	X		X		X		X		X		X		X	X	X		X		X		X		X		X		X		X		X

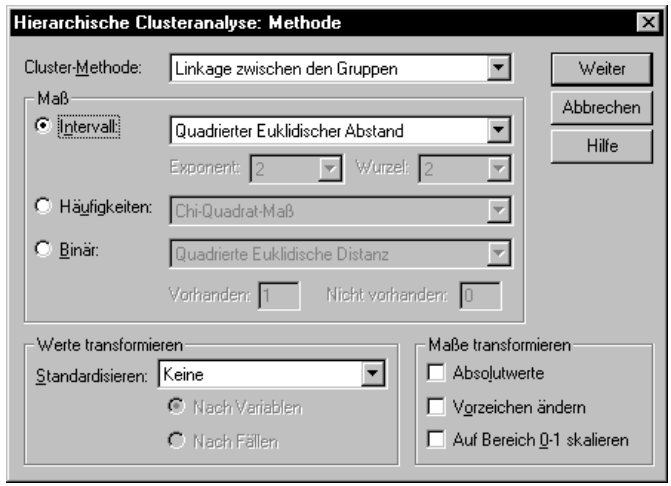
```
* * * * * * H I E R A R C H I C A L  C L U S T E R  A N A L Y S I S * * * * * *

 Dendrogram using Average Linkage (Between Groups)

                        Rescaled Distance Cluster Combine

   C A S E      0         5        10        15        20        25
 Label     Num  +---------+---------+---------+---------+---------+

 LIFEEXPF    2  ⇩×⇩⇩⇩⇩⇩⇩⇩
 BABYMORT    5  ⇩⇦      ▫⇩⇩⇩⇩⇩
 LITERACY    3  ⇩⇩⇩⇩⇩⇩⇩⇦   ▫⇩⇩⇩⇩⇩⇩⇩⇩⇩⇩⇩⇩⇩⇩⇩
 BIRTH_RT    6  ⇩×⇩⇩⇩⇩⇩⇩⇩⇩⇩⇦            ▫⇩⇩⇩⇩⇩⇩⇩⇩⇩⇩⇩⇩⇩⇩⇩
```

- Beim Clustern von Fällen müssen Sie mindestens eine numerische Variable auswählen. Beim Clustern von Variablen müssen Sie mindestens drei numerische Variablen auswählen.

  Sie haben auch die Möglichkeit, eine Variable für die Beschriftung der Fälle auszuwählen.

Hierarchische Clusteranalyse: Methode

Abbildung 33-3
Dialogfeld "Hierarchische Clusteranalyse: Methode"



Cluster-Methode. Verfügbar sind Linkage zwischen den Gruppen, Linkage innerhalb der Gruppen, nächstgelegener Nachbar, entferntester Nachbar, Zentroid-Clustering, Median-Clustering und die Ward-Methode.

Maß. Hiermit können Sie das Distanz- oder Ähnlichkeitsmaß bestimmen, das beim Clustern verwendet wird. Wählen Sie den Typ der Daten sowie das geeignete Distanz- oder Ähnlichkeitsmaß aus.

- **Intervall.** Verfügbar sind euklidische Distanz, quadrierte euklidische Distanz, Kosinus, Pearson-Korrelation, Tschebyscheff, Block, Minkowski und die Option Benutzerdefiniert.

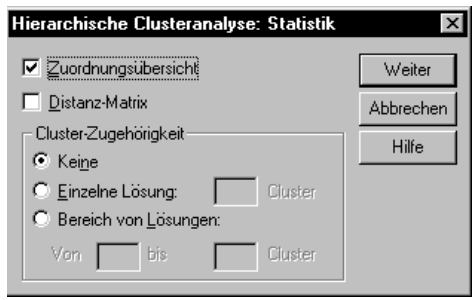
- **Häufigkeiten.** Verfügbar sind Chi-Quadratmaß und Phi-Quadratmaß.
- **Binär.** Verfügbar sind euklidische Distanz, quadrierte euklidische Distanz, Größendifferenz, Musterdifferenz, Varianz, Streuung, Form, einfache Übereinstimmung, Phi-4-Punkt-Korrelation, Lambda, Anderberg-*D*, Würfel, Hamann, Jaccard, Kulczynski 1, Kulczynski 2, Distanzmaß nach Lance und Williams, Ochiai, Ähnlichkeitsmaß nach Rogers und Tanimoto, Russel und Rao, Ähnlichkeitsmaße nach Sokal und Sneath 1 bis 5, Yule-*Y* und Yule-*Q*.

Werte transformieren. Hier können Sie festlegen, ob die Datenwerte für Fälle oder Werte vor dem Berechnen von Ähnlichkeiten standardisiert werden (nicht für binäre Daten verfügbar). Die verfügbaren Standardisierungsmethoden sind *Z*-Werte, Bereich –1 bis 1, Bereich 0 bis 1, maximale Größe von 1, Mittelwert gleich 1 und Standardabweichung gleich 1.

Maße transformieren. Hier können Sie festlegen, ob die durch das Distanzmaß erzeugten Werte transformiert werden. Dies erfolgt, nachdem das Distanzmaß berechnet wurde. Zu den verfügbaren Alternativen zählen Absolutwerte, Ändern des Vorzeichens und Skalieren auf den Bereich 0–1.

Hierarchische Clusteranalyse: Statistik

Abbildung 33-4
Dialogfeld "Hierarchische Clusteranalyse: Statistik"



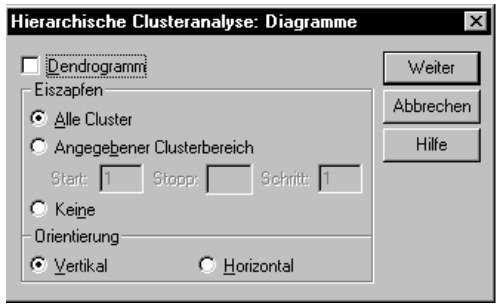
Zuordnungsübersicht. Hier wird folgendes angezeigt: Welche Fälle bzw. Cluster in jedem Schritt kombiniert wurden, die Abstände zwischen den Fällen oder Clustern, die kombiniert werden, und der Cluster-Schritt, in dem ein Fall (oder eine Variable) in den Cluster aufgenommen wurde.

Distanz-Matrix. Zeigt die Distanzen oder Ähnlichkeiten zwischen den Objekten.

Cluster-Zugehörigkeit. Zeigt den Cluster an, dem alle Fälle beim Kombinieren der Cluster in einem oder mehreren Schritten zugeordnet wurden. Die Optionen "Einzelne Lösung" und "Bereich von Lösungen" stehen zur Verfügung.

Hierarchische Clusteranalyse: Diagramme

Abbildung 33-5
Dialogfeld "Hierarchische Clusteranalyse: Grafiken"

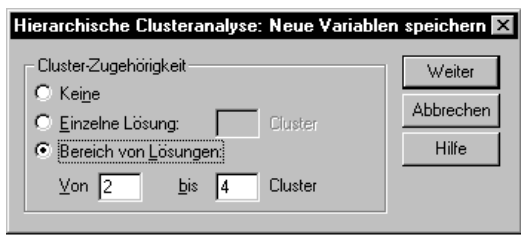


Dendrogramm. Zeigt ein **Dendrogramm**. Dendrogramme können verwendet werden, um die Dichte der gebildeten Cluster zu bewerten. Sie enthalten Informationen über die angemessene Anzahl der Cluster, die beibehalten werden sollen.

Eiszapfen. Zeigt ein **Eiszapfendiagramm**, das alle Cluster oder den angegebenen Clusterbereich enthält. Eiszapfendiagramme zeigen an, wie Fälle bei jeder Iteration der Analyse in Clustern zusammengeführt werden. Unter Orientierung können Sie ein vertikales oder horizontales Diagramm auswählen.

Hierarchische Clusteranalyse: Neue Variablen

Abbildung 33-6
Dialogfeld "Hierarchische Clusteranalyse: Neue Variablen speichern"



Cluster-Zugehörigkeit. Hiermit können Sie die Cluster-Zugehörigkeit für eine einzelne Lösung oder einen Bereich von Lösungen speichern. Die gespeicherten Variablen können dann in nachfolgenden Analysen verwendet werden, um andere Differenzen zwischen Gruppen zu untersuchen.

Zusätzliche Funktionen beim Befehl CLUSTER

In der Prozedur "Hierarchische Clusteranalyse" wird die Befehlssyntax von `CLUSTER` verwendet. Mit der SPSS-Befehlssprache verfügen Sie über die folgenden zusätzlichen Möglichkeiten:

- Verwenden mehrerer Cluster-Methoden in einer einzigen Analyse.
- Einlesen und Analysieren einer Distanzmatrix.
- Schreiben einer Distanzmatrix auf die Festplatte für eine spätere Analyse.
- Angeben aller Werte für den Exponenten und die Wurzel im benutzerdefinierten (exponentiellen) Distanzmaß.
- Festlegen der Namen für gespeicherte Variablen.

Vollständige Informationen zur Syntax finden Sie in der *SPSS Command Syntax Reference*.

Kapitel
34

Clusterzentrenanalyse

Diese Prozedur kann relativ homogene Fallgruppen aufgrund ausgewählter Eigenschaften identifizieren, wobei ein Algorithmus verwendet wird, der eine große Anzahl von Fällen verarbeiten kann. Der Algorithmus erfordert jedoch, daß Sie die Anzahl der Cluster festlegen. Wenn Ihnen die anfänglichen Clusterzentren bekannt sind, können Sie diese angeben. Sie können eine der beiden Methoden zur Klassifikation der Fälle auswählen, entweder iteratives Aktualisieren der Clusterzentren oder nur Klassifizieren. Sie können Cluster-Zugehörigkeit, Informationen zur Distanz und endgültige Clusterzentren speichern. Wahlweise können Sie eine Variable festlegen, mit deren Werte fallweise Ausgaben beschriftet werden. Sie können außerdem eine *F*-Statistik zur Varianzanalyse anfordern. Während es sich bei dieser Statistik um eine opportunistische Statistik handelt (mit dieser Prozedur wird versucht, tatsächlich voneinander abweichende Gruppen zu bilden), lassen sich aus der relativen Größe der Statistik Informationen über den Beitrag jeder Variablen zu der Trennung der Gruppen gewinnen.

Beispiel. Wodurch können Gruppen von Fernseh-Shows identifiziert werden, die innerhalb jeder Gruppe ein ähnliches Publikum anziehen? Mit der Clusterzentrenanalyse könnten Sie Fernseh-Shows (Fälle) anhand der Merkmale der Zuschauer in *k* homogene Gruppen clustern. Damit lassen sich beispielsweise Marktsegmente identifizieren. Sie können außerdem Städte (Fälle) in homogene Gruppen clustern, so daß vergleichbare Städte zum Testen verschiedener Marketingstrategien ausgewählt werden können.

Statistiken. Vollständige Lösung: anfängliche Clusterzentren, ANOVA-Tabelle. Jeder Fall: Cluster-Informationen, Distanz vom Clusterzentrum.

Daten. Die Variablen müssen quantitativ sein, entweder auf dem Intervall- oder Verhältnisniveau. Wenn Ihre Variablen binär sind oder Häufigkeiten darstellen, verwenden Sie die Prozedur “Hierarchische Clusteranalyse”.

Reihenfolge der Fälle und der anfänglichen Clusterzentren. Der Standardalgorithmus zum Auswählen der anfänglichen Clusterzentren ist nicht invariant bezüglich der Fallreihenfolge. Mit der Option Gleitende Mittelwerte verwenden im Dialogfeld "Iterieren" wird die resultierende Lösung potenziell abhängig von der Reihenfolge der Fälle, unabhängig davon, auf welche Weise die anfänglichen Clusterzentren ausgewählt wurden. Wenn Sie eine dieser Methoden nutzen, prüfen Sie daher die Stabilität einer bestimmten Lösung, indem Sie verschiedene Lösungen abrufen, bei denen die Fälle in einer unterschiedlichen, zufällig ausgewählten Reihenfolge sortiert sind. Wenn Sie anfängliche Clusterzentren angeben und dabei nicht die Option Gleitende Mittelwerte verwenden aktivieren, vermeiden Sie so potentielle Probleme im Zusammenhang mit der Fallreihenfolge. Die Reihenfolge der anfänglichen Clusterzentren kann sich jedoch auf die Lösung auswirken, wenn gebundene Distanzen von Fällen zu Clusterzentren vorliegen. Um die Stabilität einer bestimmten Lösung zu bewerten, können Sie die Ergebnisse von Analysen mit verschiedenen Permutationen der Zentrumsanfangswerte vergleichen.

Annahmen. Distanzen werden unter Verwendung des einfachen euklidischen Abstands berechnet. Wenn Sie ein anderes Distanz- oder Ähnlichkeitsmaß verwenden möchten, verwenden Sie die Prozedur "Hierarchische Clusteranalyse". Die Skalierung der Variablen ist eine wichtige Überlegung. Wenn Ihre Variablen auf unterschiedlichen Skalen gemessen wurden (wenn zum Beispiel eine Variable in Dollar und eine andere in Jahren ausgedrückt wird), können die Ergebnisse irreführend sein. In solchen Fällen sollten Sie eine Standardisierung Ihrer Variablen in Betracht ziehen, bevor Sie die Clusterzentrenanalyse durchführen (mit der Prozedur "Deskriptive Statistiken"). Diese Prozedur setzt voraus, daß Sie die passende Anzahl von Clustern ausgewählt und alle relevanten Variablen eingeschlossen haben. Wenn Sie eine ungeeignete Anzahl von Clustern ausgewählt oder wichtige Variablen ausgelassen haben, können Ihre Ergebnisse irreführend sein.

Abbildung 34-1
Ausgabe der Clusterzentrenanalyse

Anfängliche Clusterzentren

	Cluster			
	1	2	3	4
ZURBAN	-1,88606	-1,54314	1,45741	,55724
ZLIFEEXP	-3,52581	-1,69358	,62725	,99370
ZLITERAC	-2,89320	-1,65146	-,51770	,88601
ZPOP_INC	,93737	,16291	3,03701	-1,12785
ZBABYMOR	4,16813	1,38422	-,69589	-,88983
ZBIRTH_R	2,68796	,42699	,33278	-1,08033
ZDEATH_R	4,41517	,63185	-1,89037	,63185
ZLOG_GDP	-1,99641	-1,78455	,53091	1,22118
ZB_TO_D	-,52182	-,31333	4,40082	-,99285
ZFERTILT	2,24070	,75481	,46008	-,76793
ZLOG_POP	,24626	2,65246	-1,29624	-,74406

Iterationsprotokoll

		Änderung in Clusterzentren			
		1	2	3	4
Iteration	1	1.932	2.724	3.343	1.596
	2	.000	.471	.466	.314
	3	.861	.414	.172	.195
	4	.604	.337	.000	.150
	5	.000	.253	.237	.167
	6	.000	.199	.287	.071
	7	.623	.160	.000	.000
	8	.000	.084	.000	.074
	9	.000	.080	.000	.077
	10	.000	.097	.185	.000

Clusterzentren der endgültigen Lösung

	Cluster			
	1	2	3	4
ZURBAN	-1.70745	-.30863	.16816	.62767
ZLEBERW	-2.52826	-.15939	-.28417	.80611
ZLESEN	-2.30833	.13880	-.81671	.73368
ZBEV_WAC	.59747	.13400	1.45301	-.95175
ZKINDSTE	2.43210	.22286	.25622	-.80817
ZGEB_R	1.52607	.12929	1.13716	-.99285
ZSTERB_R	2.10314	-.44640	-.71414	.31319
ZLOG_BIP	-1.77704	-.58745	-.16871	.94249
ZG_ZU_V	-.29856	.19154	1.45251	-.84758
ZFRUCHT	1.51003	-.12150	1.27010	-.87669
ZLOG_EIN	.83475	.34577	-.49499	-.22199

Distanz zwischen Clusterzentren der endgültigen Lösung

		1	2	3	4
Cluster	1		5.627	5.640	7.924
	2	5.627		2.897	3.249
	3	5.640	2.897		5.246
	4	7.924	3.249	5.246	

ANOVA

	Cluster		Fehler			
	Mittel der Quadrate	df	Mittel der Quadrate	df	F	Sig.
ZURBAN	10.409	3	.541	68	19.234	.000
ZLEBERW	19.410	3	.210	68	92.614	.000
ZLESEN	18.731	3	.229	68	81.655	.000
ZBEV_WAC	18.464	3	.219	68	84.428	.000
ZKINDSTE	18.621	3	.239	68	77.859	.000
ZGEB_R	19.599	3	.167	68	117.339	.000
ZSTERB_R	13.628	3	.444	68	30.676	.000
ZLOG_BIP	17.599	3	.287	68	61.313	.000
ZG_ZU_V	16.316	3	.288	68	56.682	.000
ZFRUCHT	18.829	3	.168	68	112.273	.000
ZLOG_EIN	3.907	3	.877	68	4.457	.006

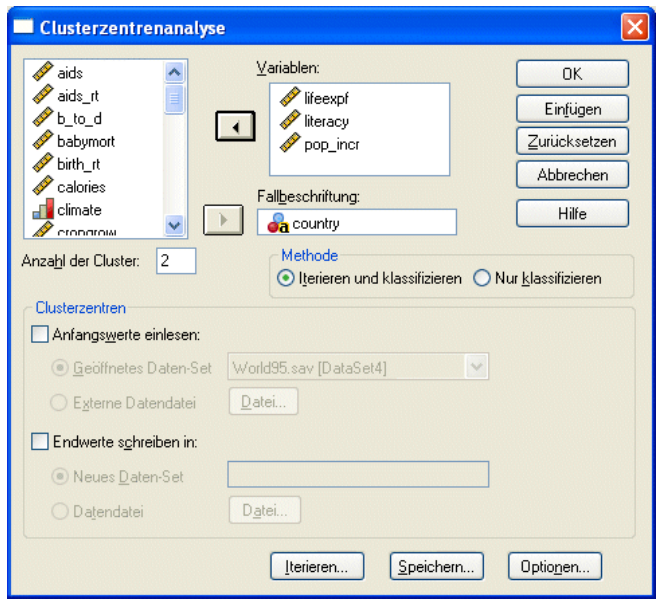
Die F-Tests sollten nur für beschreibende Zwecke verwendet werden, da die Cluster so gewählt wurden, daß die Differenzen zwischen Fällen in unterschiedlichen Clustern maximiert werden. Dabei werden die beobachteten Signifikanzniveaus nicht korrigiert und können daher nicht als Tests für die Hypothese der Gleichheit der Clustermittelwerte interpretiert werden.

So lassen Sie eine Clusterzentrenanalyse berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
Klassifizieren
Clusterzentrenanalyse…

Abbildung 34-2
Dialogfeld "Clusterzentrenanalyse"



- Wählen Sie die Variablen für die Clusteranalyse aus.
- Legen Sie die Anzahl der Cluster fest. (Die Anzahl der Cluster muss mindestens 2 betragen und darf nicht größer als die Anzahl der Fälle in der Datendatei sein.)
- Wählen Sie als Methode entweder Iterieren und klassifizieren oder Nur klassifizieren.
- Wählen Sie optional eine Identifizierungsvariable zum Beschriften der Fälle aus.

Clusterzentrenanalyse: Effizienz

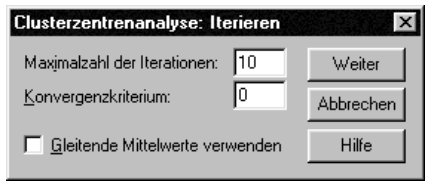
Der Befehl "Clusterzentrenanalyse" ist in erster Linie deshalb so effizient, weil er nicht die Distanzen zwischen allen Paaren von Fällen berechnet. Dies wird in vielen Algorithmen zum Clustern, auch beim hierarchischen Clustern, durchgeführt.

Für größtmögliche Effizienz nehmen Sie eine Stichprobe von Fällen und bestimmen die Clusterzentren mit der Methode Iterieren und klassifizieren. Wählen Sie Endwerte schreiben in aus. Stellen Sie anschließend die gesamte Datendatei wieder her, und wählen Sie als Methode Nur klassifizieren aus. Wählen Sie Anfangswerte

einlesen, um die gesamte Datei anhand der aus der Stichprobe geschätzten Clusterzentren zu klassifizieren. Die Daten können in eine Datei oder in ein Daten-Set geschrieben und aus einer Datei oder einem Daten-Set ausgelesen werden. Daten-Sets sind für die anschließende Verwendung in der gleichen Sitzung verfügbar, werden jedoch nicht als Dateien gespeichert, sofern Sie diese nicht ausdrücklich vor dem Beenden der Sitzung speichern. Die Namen von Daten-Sets müssen den SPSS-Regeln zum Benennen von Variablen entsprechen. Für weitere Informationen siehe “Variablennamen” in Kapitel 5 auf S. 99.

Clusterzentrenanalyse: Iterieren

Abbildung 34-3
Dialogfeld “Clusterzentrenanalyse: Iterieren”



Anmerkung: Diese Optionen sind nur verfügbar, wenn Sie im Dialogfeld “Clusterzentrenanalyse” die Methode Iterieren und klassifizieren auswählen.

Anzahl der Iterationen. Begrenzt die Anzahl der Iterationen im Clusterzentren-Algorithmus. Die Iteration wird nach der vorgegebenen Anzahl der Iterationen beendet, auch wenn das Konvergenzkriterium noch nicht erreicht wurde. Diese Zahl muß zwischen 1 und 999 liegen.

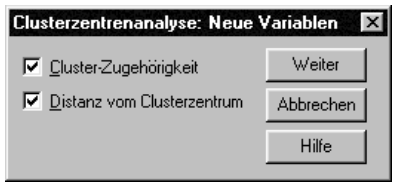
Um den Algorithmus zu verwenden, der beim Befehl “Quick Cluster” in SPSS-Versionen vor Version 5.0 verwendet wurde, setzen Sie Anzahl der Iterationen auf 1.

Konvergenzkriterium. Bestimmt, wann die Iteration beendet ist. Das Konvergenzkriterium gibt einen Anteil der minimalen Distanz zwischen anfänglichen Clusterzentren wieder. Der Wert muss also größer als 0, darf aber nicht größer als 1 sein. Wenn das Kriterium zum Beispiel 0,02 lautet, ist die Iteration beendet, sobald eine vollständige Iteration keines der Clusterzentren um eine Distanz von mehr als 2 % der kleinsten Distanz zwischen beliebigen anfänglichen Clusterzentren bewegt.

Gleitende Mittelwerte verwenden. Mit dieser Funktion können Sie eine Aktualisierung der Clusterzentren veranlassen, nachdem jeder Fall zugeordnet wurde. Wenn Sie diese Option nicht auswählen, werden neue Clusterzentren berechnet, nachdem alle Fälle zugeordnet wurden.

Clusterzentrenanalyse: Neue Variablen

Abbildung 34-4
Dialogfeld "Clusterzentrenanalyse: Neue Variablen"



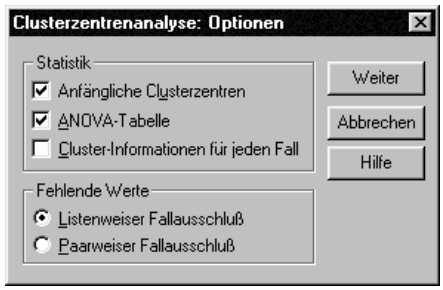
Sie können die Informationen über die Lösung als neue Variablen speichern, um diese in nachfolgenden Analysen zu verwenden:

Cluster-Zugehörigkeit. Erstellt eine neue Variable, welche die endgültige Cluster-Zugehörigkeit für jeden Fall anzeigt. Die Werte der neuen Variablen liegen in einem Bereich von 1 bis zur Anzahl der Cluster.

Distanz vom Clusterzentrum. Erstellt eine neue Variable, welche den euklidischen Abstand zwischen jedem Fall und seinem Klassifikationszentrum anzeigt.

Clusterzentrenanalyse: Optionen

Abbildung 34-5
Dialogfeld "Clusterzentrenanalyse: Optionen"



Statistiken. Sie können die folgenden Statistiken auswählen: anfängliche Clusterzentren, ANOVA-Tabelle und Cluster-Information für jeden Fall.

- **Anfängliche Clusterzentren.** Erste Schätzung der Mittelwerte der Variablen für jeden Cluster. In der Standardeinstellung werden zunächst so viele günstig gelegene Fälle aus den Daten ausgewählt, wie Cluster gebildet werden sollen. Die anfänglichen Clusterzentren werden für eine Ausgangsklassifizierung verwendet und dann aktualisiert.
- **ANOVA-Tabelle (K-Means Cluster Options).** Zeigt eine Varianzanalysetabelle mit univariaten F-Tests für jede Cluster-Variable an. Die F-Tests haben nur beschreibenden Charakter, und die daraus resultierenden Wahrscheinlichkeiten sind nicht zu interpretieren. Die ANOVA-Tabelle wird nicht angezeigt, wenn alle Fälle einem einzigen Cluster zugewiesen werden.
- **Clusterinformationen für die einzelnen Fälle.** Zeigt für jeden Fall die endgültige Clusterzuordnung und den Euklidischen Abstand zwischen dem Fall und dem Clusterzentrum, das zur Klassifizierung des Falles verwendet wird. Es werden auch die euklidischen Abstände zwischen den endgültigen Clusterzentren angezeigt.

Fehlende Werte. Die verfügbaren Optionen sind Listenweiser Fallausschluß oder Paarweiser Fallausschluß.

- **Listenweiser Fallausschluß.** Fälle, bei denen Werte einer beliebigen Clustervariable fehlen, werden aus der Analyse ausgeschlossen.
- **Paarweiser Fallausschluß.** Die Fälle werden den Clustern auf der Grundlage der aus allen Variablen mit nichtfehlenden Werten berechneten Distanzen zugewiesen.

Zusätzliche Funktionen beim Befehl QUICK CLUSTER

In der Prozedur "Clusterzentrenanalyse" wird die Befehlssyntax von `QUICK CLUSTER` verwendet. Mit der SPSS-Befehlssprache verfügen Sie über die folgenden zusätzlichen Möglichkeiten:

- Übernehmen der ersten *k* Fälle als anfängliche Clusterzentren. Dadurch wird der üblicherweise für deren Schätzung benötigte Verarbeitungsdurchlauf vermieden.
- Direktes Angeben der anfänglichen Clusterzentren als Teil der Befehlssyntax.
- Festlegen der Namen für gespeicherte Variablen.

Vollständige Informationen zur Syntax finden Sie in der *SPSS Command Syntax Reference*.

Kapitel
35

Nichtparametrische Tests

Mit der Prozedur "Nichtparametrische Tests" stehen Ihnen verschiedene Tests zur Verfügung, bei denen keine Annahmen über die Form der zugrunde liegenden Verteilung benötigt werden.

Chi-Quadrat-Test. Mit diesem Test wird eine Variable nach Kategorien aufgelistet und auf der Grundlage der Differenzen zwischen beobachteten und erwarteten Häufigkeiten eine Chi-Quadrat-Statistik berechnet.

Test auf Binomialverteilung. In diesem Test wird die beobachtete Häufigkeit in jeder Kategorie einer dichotomen Variablen mit den erwarteten Häufigkeiten der binomialen Verteilung verglichen.

Sequenztest. Hiermit können Sie testen, ob zwei Werte einer Variablen in zufälliger Reihenfolge auftreten.

Kolmogorov-Smirnov-Test bei einer Stichprobe. Hierbei wird die beobachtete kumulative Verteilungsfunktion einer Variablen mit einer bestimmten theoretischen Verteilung verglichen. Bei der Verteilung kann es sich um eine Normalverteilung, eine Gleichverteilung, Exponentialverteilung oder um eine Poisson-Verteilung handeln.

Test bei zwei unabhängigen Stichproben. Mit diesem Test können zwei Fallgruppen bei einer Variablen verglichen werden. Dabei stehen die folgenden Tests zur Verfügung: Mann-Whitney-*U*-Test, Kolmogorov-Smirnov-Test bei zwei Stichproben, Test auf Extremreaktionen nach Moses und Sequenzentest nach Wald-Wolfowitz.

Tests bei zwei verbundenen Stichproben. Hiermit können die Verteilungen von zwei Variablen verglichen werden. Dafür stehen der Wilcoxon-Test, der Vorzeichentest und der McNemar-Test zur Verfügung.

Test bei mehreren unabhängigen Stichproben. Hiermit können Sie zwei oder mehrere Fallgruppen bei einer Variablen vergleichen. Dafür stehen der Kruskal-Wallis-H-Test, der Mediantest und der Jonckheere-Terpstra-Test zur Verfügung.

Tests bei mehreren verbundenen Stichproben. Hiermit können Sie die Verteilungen von zwei oder mehr Variablen vergleichen. Dafür stehen der Friedman-Test, Kendall-*W* und Cochrans *Q*-Test zur Verfügung.

Bei allen oben aufgeführten Tests können Quartile, Mittelwert, Standardabweichung, Minimum, Maximum und die Anzahl nichtfehlender Fälle berechnet werden.

Chi-Quadrat-Test

Mit der Prozedur "Chi-Quadrat-Test" können Sie eine Variable nach Kategorien auflisten und eine Chi-Quadrat-Statistik berechnen lassen. Bei diesem Anpassungstest werden die beobachteten und erwarteten Häufigkeiten in allen Kategorien miteinander verglichen. Dadurch wird überprüft, ob entweder alle Kategorien den gleichen Anteil an Werten enthalten oder ob jede Kategorie jeweils einen vom Benutzer festgelegten Anteil an Werten enthält.

Beispiel. Mithilfe des Chi-Quadrat-Tests können Sie bestimmen, ob in einer Tüte mit Gummibärchen die gleiche Anzahl an weißen, grünen, orangefarbenen, roten und gelben Gummibärchen vorhanden sind. Sie können auch prüfen, ob eine Tüte 30% weiße, 17% grüne, 23% orangefarbene, 15% rote und 15% gelbe Gummibärchen enthält.

Statistiken. Mittelwert, Standardabweichung, Minimum, Maximum und Quartile. Die Anzahl und der Prozentsatz nichtfehlender und fehlender Fälle, die Anzahl der für jede Kategorie beobachteten und erwarteten Fälle, Residuen und die Chi-Quadrat-Statistik.

Daten. Verwenden Sie geordnete oder nichtgeordnete numerische kategoriale Variablen (nominales oder ordinales Niveau der Meßwerte). Verwenden Sie zum Umwandeln von String-Variablen in numerische Variablen den Befehl "Automatisch umkodieren" im Menü "Transformieren".

Annahmen. Nichtparametrische Tests erfordern keine Annahmen über die Form der zugrundeliegenden Verteilung. Die Daten werden als zufällige Stichprobe betrachtet. Die erwartete Häufigkeit in jeder Kategorie muß mindestens 1 betragen. Bei höchstens 20% der Kategorien darf die erwartete Häufigkeit unter 5 liegen.

Abbildung 35-1
Ausgabe für den Chi-Quadrat-Test

Farbe der Gummibärchen

	Beobachtetes N	Erwartete Anzahl	Residuum
Blau	6	18.8	-12.8
Braun	33	18.8	14.2
Grün	9	18.8	-9.8
Gelb	17	18.8	-1.8
Orange	22	18.8	3.2
Rot	26	18.8	7.2
Gesamt	113		

Statistik für Test

	Farbe der Gummibärchen
Chi-Quadrat[1]	27.973
df	5
Asymptotische Signifikanz	.000

1. Bei 0 Zellen (.0%) werden weniger als 5 Häufigkeiten erwartet. Die kleinste erwartete Zellenhäufigkeit ist 18.8.

Farbe der Gummibärchen

	Beobachtetes N	Erwartete Anzahl	Residuum
Blau	6	5.9	.1
Braun	33	35.7	-2.7
Grün	9	11.9	-2.9
Gelb	17	23.8	-6.8
Orange	22	17.8	4.2
Rot	26	17.8	8.2
Gesamt	113		

Statistik für Test

	Farbe der Gummibärchen
Chi-Quadrat[1]	7.544
df	5
Asymptotische Signifikanz	.183

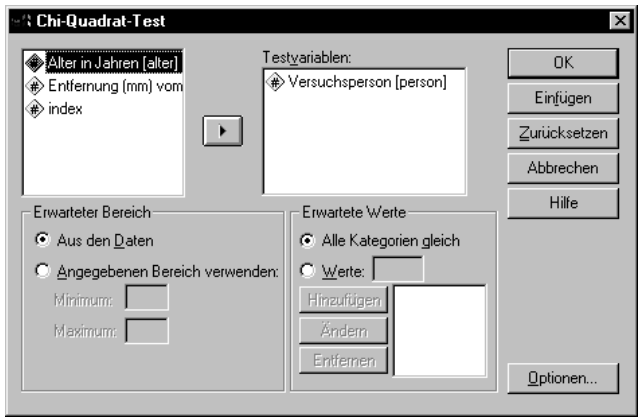
1. Bei 0 Zellen (.0%) werden weniger als 5 Häufigkeiten erwartet. Die kleinste erwartete Zellenhäufigkeit ist 5.9.

So lassen Sie einen Chi-Quadrat-Test berechnen

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Nichtparametrische Tests
  Chi-Quadrat

Abbildung 35-2
Dialogfeld "Chi-Quadrat-Test"



- Wählen Sie mindestens eine Testvariable aus. Mit jeder Variablen wird ein separater Test erzeugt.

- Wenn Sie auf Optionen klicken, können Sie deskriptive Statistiken und Quartile abrufen sowie festlegen, wie fehlende Werte verarbeitet werden.

Chi-Quadrat-Test: erwarteter Bereich und erwartete Werte

Erwarteter Bereich. In der Standardeinstellung wird jeder einzelne Wert einer Variablen als eine Kategorie definiert. Zum Aufstellen von Kategorien in einem bestimmten Bereich wählen Sie AngegebenenBereich verwenden, und geben Sie für die obere und die untere Grenze jeweils einen ganzzahligen Wert an. Für jeden ganzzahligen Wert in dem eingeschlossenen Bereich wird eine Kategorie aufgestellt, wobei Fälle mit Werten außerhalb der angegebenen Grenzen ausgeschlossen werden. Wenn Sie zum Beispiel für das Minimum den Wert 1 und für das Maximum den Wert 4 angeben, werden für den Chi-Quadrat-Test nur die Werte von 1 bis 4 verwendet.

Erwartete Werte. In der Standardeinstellung sind die erwarteten Werte für alle Kategorien gleich. Die erwarteten Anteile der Kategorien können vom Benutzer festgelegt werden. Wählen Sie Werte aus. Geben Sie für jede Kategorie der Testvariablen einen Wert größer als 0 ein, und klicken Sie dann auf Hinzufügen. Jeder neu eingegebene Wert wird am Ende der Werteliste angezeigt. Die Reihenfolge der Werte ist von Bedeutung. Sie entspricht der aufsteigenden Folge der Kategoriewerte für die Testvariable. Der erste Wert in der Liste entspricht dem niedrigsten Gruppenwert der Testvariablen, der letzte Wert entspricht dem höchsten Wert. Die Elemente der Werteliste werden summiert. Anschließend wird jeder Wert durch diese Summe dividiert, um den Anteil der in der entsprechenden Kategorie erwarteten Fälle zu berechnen. So ergibt eine Werteliste mit 3, 4, 5 und 4 beispielsweise die erwarteten Anteile 3/16, 4/16, 5/16 und 4/16.

Chi-Quadrat-Test: Optionen

Abbildung 35-3
Dialogfeld "Chi-Quadrat-Test: Optionen"

Statistik. Sie können eine oder beide Auswertungsstatistiken wählen.

- **Deskriptive Statistik.** Bei dieser Option werden Mittelwert, Standardabweichung, Minimum, Maximum und Anzahl der nichtfehlenden Fälle angezeigt.
- **Quartile.** Zeigt die Werte an, die den 25., 50. und 75. Perzentilen entsprechen.

Fehlende Werte. Bestimmt die Verarbeitung fehlender Werte.

- **Fallausschluß Test für Test.** Werden mehrere Tests festgelegt, so wird jeder Test einzeln auf fehlende Werte geprüft.
- **Listenweiser Fallausschluß.** Fälle mit fehlenden Werten für eine Variable werden aus allen Analysen ausgeschlossen.

Zusätzliche Funktionen beim Befehl NPAR TESTS (Chi-Quadrat-Test)

Mit der SPSS-Befehlssprache verfügen Sie über die folgenden zusätzlichen Möglichkeiten:

- Mit dem Unterbefehl `CHISQUARE` können verschiedene Minimal- und Maximalwerte sowie erwartete Häufigkeiten für verschiedene Variablen angegeben werden.
- Mit dem Unterbefehl `EXPECTED` kann eine Variable bei verschiedenen erwarteten Häufigkeiten getestet werden, oder es können verschiedene Bereiche verwendet werden.

Vollständige Informationen zur Syntax finden Sie in der *SPSS Command Syntax Reference*.

Test auf Binomialverteilung

Mit der Prozedur "Test auf Binomialverteilung" können Sie die beobachteten Häufigkeiten der beiden Kategorien einer dichotomen Variablen mit den Häufigkeiten vergleichen, die unter einer Binomialverteilung mit einem angegebenen Wahrscheinlichkeitsparameter zu erwarten sind. In der Standardeinstellung ist der Wahrscheinlichkeitsparameter für beide Gruppen auf 0,5 gesetzt. Zum Ändern der Wahrscheinlichkeiten können Sie einen Testanteil für die erste Gruppe angeben. Die Wahrscheinlichkeit für die zweite Gruppe beträgt 1 minus der für die erste Gruppe angegebenen Wahrscheinlichkeit.

Beispiel. Wenn Sie eine Münze werfen, ist die Wahrscheinlichkeit, daß diese mit dem Kopf nach oben zu liegen kommt, gleich 1/2. Auf der Grundlage dieser Hypothese wird nun eine Münze 40mal geworfen, wobei die Ergebnisse aufgezeichnet werden (Kopf oder Zahl). Der Test auf Binomialverteilung könnte dann beispielsweise ergeben, daß 3/4 der Würfe "Kopf" waren und das beobachtete Signifikanzniveau gering ist (0,0027). Diese Ergebnisse zeigen an, daß die Wahrscheinlichkeit für "Kopf" nicht 1/2 beträgt und die Münze somit wahrscheinlich manipuliert ist.

Statistiken. Mittelwert, Standardabweichung, Minimum, Maximum, Anzahl der nichtfehlenden Fälle und Quartile.

Daten. Die getesteten Variablen müssen numerisch und dichotom sein. Verwenden Sie zum Umwandeln von String-Variablen in numerische Variablen den Befehl "Automatisch umkodieren" im Menü "Transformieren". **Dichotome Variablen** sind

Variablen, die nur zwei mögliche Werte annehmen können: *ja* oder *nein*, *wahr* oder *falsch*, 0 oder 1 usw. Wenn die Varaiblen nicht dichotom sind, müssen Sie einen Trennwert angeben. Durch den Trennwert werden Fälle mit Werten über dem Trennwert einer Gruppe und alle anderen Fälle einer anderen Gruppe zugeordnet.

Annahmen. Nichtparametrische Tests erfordern keine Annahmen über die Form der zugrundeliegenden Verteilung. Die Daten werden als zufällige Stichprobe betrachtet.

Abbildung 35-4
Ausgabe des Tests auf Binomialverteilung

Korrelationen

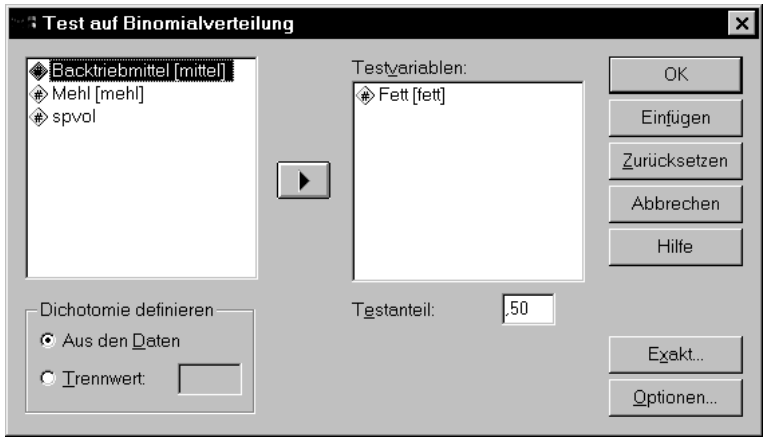
		Gewonnene Spiele	Eigene Punkte pro Spiel	Gegnerische Punkte pro Spiel
Korrelation nach Pearson	Gewonnene Spiele	1.000	.581	-.401
	Eigene Punkte pro Spiel	.581	1.000	.457
	Gegnerische Punkte pro Spiel	-.401	.457	1.000
Signifikanz (2-seitig)	Gewonnene Spiele	.	.001	.038
	Eigene Punkte pro Spiel	.001	.	.017
	Gegnerische Punkte pro Spiel	.038	.017	.
N	Gewonnene Spiele	27	27	27
	Eigene Punkte pro Spiel	27	27	27
	Gegnerische Punkte pro Spiel	27	27	27

So lassen Sie einen Test auf Binomialverteilung berechnen

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Nichtparametrische Tests
    Binomial…

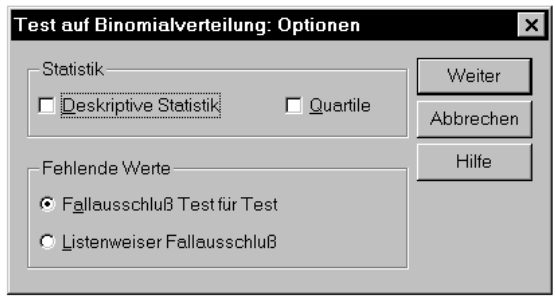
Abbildung 35-5
Dialogfeld "Test auf Binomialverteilung"



- Wählen Sie mindestens eine numerische Testvariable.
- Wenn Sie auf Optionen klicken, können Sie deskriptive Statistiken und Quartile abrufen sowie festlegen, wie fehlende Werte verarbeitet werden.

Optionen für den Test auf Binomialverteilung

Abbildung 35-6
Dialogfeld "Test auf Binomialverteilung: Optionen"



Statistik. Sie können eine oder beide Auswertungsstatistiken wählen.

- **Deskriptive Statistik.** Bei dieser Option werden Mittelwert, Standardabweichung, Minimum, Maximum und Anzahl der nichtfehlenden Fälle angezeigt.
- **Quartile.** Zeigt die Werte an, die den 25., 50. und 75. Perzentilen entsprechen.

Fehlende Werte. Bestimmt die Verarbeitung fehlender Werte.

- **Fallausschluß Test für Test.** Werden mehrere Tests festgelegt, so wird jeder Test einzeln auf fehlende Werte geprüft.
- **Listenweiser Fallausschluß.** Fälle mit fehlenden Werten für eine beliebige getestete Variable werden von allen Analysen ausgeschlossen.

Zusätzliche Funktionen beim Befehl NPAR TESTS (Test auf Binomialverteilung)

Mit der SPSS-Befehlssprache verfügen Sie über die folgenden zusätzlichen Möglichkeiten:

- Mit dem Unterbefehl `BINOMIAL` können bestimmte Gruppen ausgewählt und andere Gruppen ausgeschlossen werden, wenn eine Variable über mehr als zwei Kategorien verfügt.
- Mit dem Unterbefehl `BINOMIAL` können verschiedene Trennwerte oder Wahrscheinlichkeiten für verschiedene Variablen angeben werden.
- Mit dem Unterbefehl `EXPECTED` kann dieselbe Variable bei verschiedenen Trennwerten oder Wahrscheinlichkeiten getestet werden.

Vollständige Informationen zur Syntax finden Sie im *SPSS Syntax Reference Guide*.

Sequenzentest

Mit der Prozedur “Sequenzentest” können Sie testen, ob zwei Werte einer Variablen in zufälliger Reihenfolge auftreten. Eine Sequenz ist eine Folge von gleichen Beobachtungen. Eine Stichprobe mit zu vielen oder zu wenigen Sequenzen legt nahe, daß die Stichprobe nicht zufällig ist.

Beispiel. Es werden 20 Personen befragt, ob sie ein bestimmtes Produkt kaufen würden. Die angenommene zufällige Auswahl der Stichprobe wäre ernsthaft zu bezweifeln, wenn alle 20 Personen demselben Geschlecht angehören würden. Mit dem Sequenzentest kann bestimmt werden, ob die Stichprobe zufällig entnommen wurde.

Statistiken. Mittelwert, Standardabweichung, Minimum, Maximum, Anzahl der nichtfehlenden Fälle und Quartile.

Daten. Die Variablen müssen numerisch sein. Verwenden Sie zum Umwandeln von String-Variablen in numerische Variablen den Befehl "Automatisch umkodieren" im Menü "Transformieren".

Annahmen. Nichtparametrische Tests erfordern keine Annahmen über die Form der zugrundeliegenden Verteilung. Verwenden Sie Stichproben aus stetigen Wahrscheinlichkeitsverteilungen.

Abbildung 35-7
Ausgabe des Sequenzentests

Runs Test

	Gender
Test Value [1]	1.00
Cases < Test Value	7
Cases >= Test Value	13
Total Cases	20
Number of Runs	15
Z	2.234
Asymptotic Significance (2-tailed)	.025

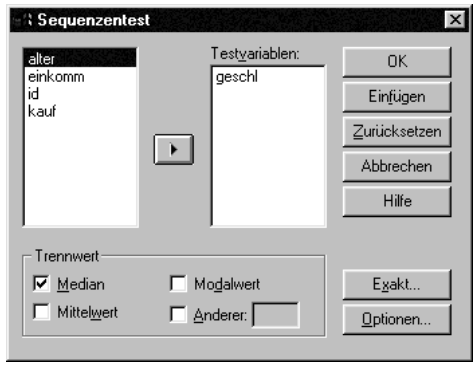
1. Median

So lassen Sie einen Sequenzentest berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Nichtparametrische Tests
    Sequenzen...

Abbildung 35-8
Dialogfeld "Sequenzentest"



- Wählen Sie mindestens eine numerische Testvariable.
- Wenn Sie auf Optionen klicken, können Sie deskriptive Statistiken und Quartile abrufen sowie festlegen, wie fehlende Werte verarbeitet werden.

Sequenzentest: Trennwert

Trennwert. Hier wird ein Trennwert zum Dichotomisieren der gewählten Variablen angegeben. Sie können den beobachteten Mittelwert, den Median, den Modalwert oder einen angegebenen Wert als Trennwert wählen. Fälle mit Werten kleiner als der Trennwert werden einer Gruppe, Fälle mit Werten größer oder gleich dem Trennwert einer anderen Gruppe zugeordnet. Für jeden gewählten Trennwert wird ein Test ausgeführt.

Sequenzentest: Optionen

Abbildung 35-9
Dialogfeld "Sequenzentest: Optionen"

Statistik. Sie können eine oder beide Auswertungsstatistiken wählen.

- **Deskriptive Statistik.** Bei dieser Option werden Mittelwert, Standardabweichung, Minimum, Maximum und Anzahl der nichtfehlenden Fälle angezeigt.
- **Quartile.** Zeigt die Werte an, die den 25., 50. und 75. Perzentilen entsprechen.

Fehlende Werte. Bestimmt die Verarbeitung fehlender Werte.

- **Fallausschluß Test für Test.** Werden mehrere Tests festgelegt, so wird jeder Test einzeln auf fehlende Werte geprüft.
- **Listenweiser Fallausschluß.** Fälle mit fehlenden Werten für eine Variable werden aus allen Analysen ausgeschlossen.

Zusätzliche Funktionen beim Befehl NPAR TESTS (Sequenzentest)

Mit der SPSS-Befehlssprache verfügen Sie über die folgenden zusätzlichen Möglichkeiten:

- Mit dem Unterbefehl `RUNS` können verschiedene Trennwerte für verschiedene Variablen angegeben werden.
- Mit dem Unterbefehl `RUNS` kann dieselbe Variable mit verschiedenen benutzerdefinierten Trennwerten getestet werden.

Vollständige Informationen zur Syntax finden Sie im *SPSS Syntax Reference Guide*.

Kolmogorov-Smirnov-Test bei einer Stichprobe

Mit dem Kolmogorov-Smirnov-Test bei einer Stichprobe (Anpassungstest) wird die beobachtete kumulative Verteilungsfunktion für eine Variable mit einer festgelegten theoretischen Verteilung verglichen, die eine Normalverteilung, eine Gleichverteilung, eine Poisson-Verteilung oder Exponentialverteilung sein kann. Das Kolmogorov-Smirnov-*Z* wird aus der größten Differenz (in Absolutwerten) zwischen beobachteten und theoretischen kumulativen Verteilungsfunktionen berechnet. Mit diesem Test für die Güte der Anpassung wird getestet, ob die Beobachtung wahrscheinlich aus der angegebenen Verteilung stammt.

Beispiel. Für viele parametrische Tests sind normalverteilte Variablen erforderlich. Mit dem Kolmogorov-Smirnov-Anpassungstest kann getestet werden, ob eine Variable, zum Beispiel *Einkommen*, normalverteilt ist.

Statistiken. Mittelwert, Standardabweichung, Minimum, Maximum, Anzahl der nichtfehlenden Fälle und Quartile.

Daten. Die Variablen müssen auf Intervall- oder Verhältnis-Meßniveau quantitativ sein.

Annahmen. Für den Kolmogorov-Smirnov-Test wird angenommen, daß die Parameter der zu testenden Verteilung im voraus angegeben wurden. Mit dieser Prozedur werden die Parameter aus der Stichprobe geschätzt. Der Mittelwert und die Standardabweichung der Stichprobe sind die Parameter für eine Normalverteilung. Minimum und Maximum der Stichprobe definieren die Spannweite der Gleichverteilung, und der Mittelwert der Stichprobe ist der Parameter für die Poisson-Verteilung sowie der Parameter für die Exponentialverteilung. Die Stärke des Tests, Abweichungen von der hypothetischen Verteilung zu erkennen, kann dabei deutlich verringert werden. Wenn Sie einen Test gegen eine Normalverteilung mit geschätzten Parametern durchführen möchten, sollten Sie den Kolmogorov-Smirnov-Test mit der Korrektur nach Lilliefors (in der Prozedur “Explorative Datenanalyse”) in Betracht ziehen.

Abbildung 35-10
Ausgabe des Kolmogorov-Smirnov-Anpassungstests

Kolmogorov-Smirnov-Anpassungstest

		Einkommen
N		20
Parameter der Normalverteilung[1,2]	Mittelwert	56250.00
	Standardabweichung	45146.40
Extremste Differenzen	Absolut	.170
	Positiv	.170
	Negativ	-.164
Kolmogorov-Smirnov-Z		.760
Asymptotische Signifikanz (2-seitig)		.611

1. Die zu testende Verteilung ist eine Normalverteilung.
2. Aus den Daten berechnet.

So berechnen Sie einen Kolmogorov-Smirnov-Anpassungstest:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Nichtparametrische Tests
    K-S bei einer Stichprobe…

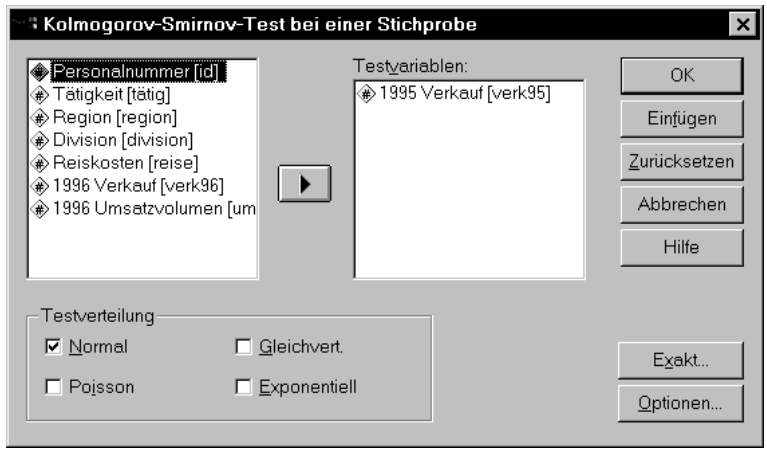
Abbildung 35-11
Dialogfeld "Kolmogorov-Smirnov-Test bei einer Stichprobe"

▶ Wählen Sie mindestens eine numerische Testvariable. Mit jeder Variablen wird ein separater Test erzeugt.

- Wenn Sie auf Optionen klicken, können Sie deskriptive Statistiken und Quartile abrufen sowie festlegen, wie fehlende Werte verarbeitet werden.

K-S bei einer Stichprobe: Optionen

Abbildung 35-12
Dialogfeld "K-S bei einer Stichprobe: Optionen"



Statistik. Sie können eine oder beide Auswertungsstatistiken wählen.

- **Deskriptive Statistik.** Bei dieser Option werden Mittelwert, Standardabweichung, Minimum, Maximum und Anzahl der nichtfehlenden Fälle angezeigt.
- **Quartile.** Zeigt die Werte an, die den 25., 50. und 75. Perzentilen entsprechen.

Fehlende Werte. Bestimmt die Verarbeitung fehlender Werte.

- **Fallausschluß Test für Test.** Werden mehrere Tests festgelegt, so wird jeder Test einzeln auf fehlende Werte geprüft.
- **Listenweiser Fallausschluß.** Fälle mit fehlenden Werten für eine Variable werden aus allen Analysen ausgeschlossen.

Zusätzliche Funktionen beim Befehl NPAR TESTS (Kolmogorov-Smirnov-Anpassungstest)

Mit der SPSS-Befehlssprache können Sie auch die Parameter der zu testenden Verteilung angeben (mit dem Unterbefehl `K-S`).

Vollständige Informationen zur Syntax finden Sie im *SPSS Syntax Reference Guide*.

Tests bei zwei unabhängigen Stichproben

Die Prozedur "Test bei zwei unabhängigen Stichproben" vergleicht zwei Gruppen von Fällen von einer Variablen.

Beispiel. Es wurden neue Zahnspangen entwickelt, die bequemer sein sollen, besser aussehen und zu einem schnelleren Erfolg beim Richten der Zähne führen sollen. Um festzustellen, ob die neuen Spangen so lange wie die alten getragen werden müssen, wurden willkürlich 10 Kinder zum Tragen der alten Zahnspangen und weitere 10 Kinder zum Tragen der neuen Spangen ausgewählt. Anhand des Mann-Whitney-*U*-Tests stellen Sie eventuell fest, dass die neuen Spangen im Durchschnitt nicht so lange wie die alten Spangen getragen werden mussten.

Statistiken. Mittelwert, Standardabweichung, Minimum, Maximum, Anzahl der nichtfehlenden Fälle und Quartile. Tests: Mann-Whitney-*U*-Test, Extremreaktionen nach Moses, Kolmogorov-Smirnov-*Z*-Test, Sequenztest nach Wald-Wolfowitz.

Daten. Verwenden Sie numerische Variablen, die geordnet werden können.

Annahmen. Verwenden Sie unabhängige Zufallsstichproben. Der Mann-Whitney-*U*-Test erfordert, dass sich die beiden getesteten Stichproben in ihrer Form ähneln.

Abbildung 35-13
Ausgabe für "Zwei unabhängige Stichproben"

Ränge

			N	Mittlerer Rang	Rangsumme
Tragezeit in Tagen	Zahnspangenmodell	Alt	10	14.10	141.00
		Neu	10	6.90	69.00
		Gesamt	20		

Statistik für Test[2]

	Tragezeit in Tagen
Mann-Whitney-U	14.000
Wilcoxon-W	69.000
Z	-2.721
Asymptotische Signifikanz (2-seitig)	.007
Exakte Signifikanz [2*(1-seitig Sig.)]	.005[1]

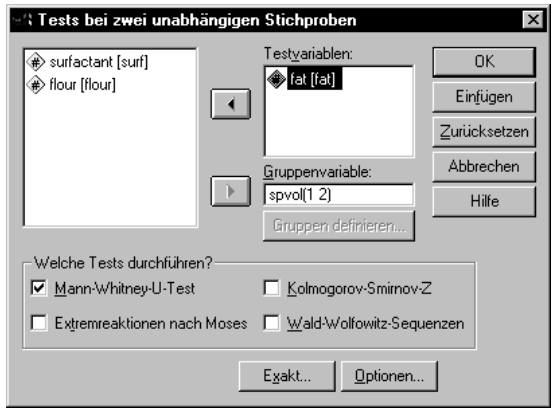
1. Nicht für Bindungen korrigiert.
2. Gruppenvariable: Zahnspangenmodell

So lassen Sie Tests bei zwei unabhängigen Stichproben berechnen:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Nichtparametrische Tests
  Zwei unabhängige Stichproben...

Abbildung 35-14
Dialogfeld "Tests bei zwei unabhängigen Stichproben"



- Wählen Sie mindestens eine numerische Variable aus.
- Wählen Sie eine Gruppenvariable aus, und klicken Sie auf Gruppen definieren, um die Datei in zwei Gruppen oder Stichproben aufzuteilen.

Typen von Tests bei zwei unabhängigen Stichproben

Welche Tests durchführen? Mit Hilfe von vier Tests können Sie überprüfen, ob zwei unabhängige Stichproben (Gruppen) aus derselben Grundgesamtheit stammen.

Der **Mann-Whitney-U-Test** ist der am häufigsten verwendete Test bei zwei unabhängigen Stichproben. Er ist äquivalent zum Wilcoxon-Rangsummentest und dem Kruskal-Wallis-Test für zwei Gruppen. Mit dem Mann-Whitney-U-Test wird überprüft, ob zwei beprobte Grundgesamtheiten die gleiche Lage besitzen. Die Beobachtungen aus beiden Gruppen werden kombiniert und in eine gemeinsame Reihenfolge gebracht, wobei im Falle von Rangbindungen der durchschnittliche Rang vergeben wird. Die Anzahl der Bindungen sollte im Verhältnis zur Gesamtanzahl der Beobachtungen klein sein. Wenn die Grundgesamtheiten in der Lage identisch sind, sollten die Ränge zufällig zwischen den beiden Stichproben gemischt werden. Im Test wird berechnet, wie oft ein Wert aus Gruppe 1 einem Wert aus Gruppe 2 und wie oft ein Wert aus Gruppe 2 einem Wert aus Gruppe 1 vorangeht. Die Mann-Whithney-*U*-Statistik ist die kleinere dieser beiden Zahlen. Die ebenfalls angezeigte Wilcoxon-Rangsummen-*W*-Statistik ist die Rangsumme der kleineren Stichprobe. Wenn beide Stichproben die gleiche Anzahl von Beobachtungen aufweisen, ist *W* die Rangsumme der zuerst benannten Gruppe im Dialogfeld "Zwei unabhängige Stichproben: Gruppen definieren".

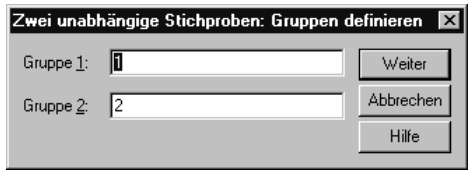
Der **Kolmogorov-Smirnov-Z-Test** und der **Sequenztest nach Wald-Wolfowitz** stellen eher allgemeine Tests dar, die sowohl Unterschiede in den Lagen als auch in den Formen der Verteilungen erkennen. Der Test nach Kolmogorov-Smirnov arbeitet auf der Grundlage der maximalen absoluten Differenz zwischen den beobachteten kumulativen Verteilungsfunktionen für beide Stichproben. Wenn diese Differenz signifikant groß ist, werden die beiden Verteilungen als verschieden betrachtet. Der Sequenztest nach Wald-Wolfowitz kombiniert die Beobachtungen aus beiden Gruppen und ordnet ihnen einen Rang zu. Wenn die beiden Stichproben aus derselben Grundgesamtheit stammen, müssen die beiden Gruppen in der Rangverteilung zufällig gestreut sein.

Der **Test "Extremreaktionen nach Moses"** setzt voraus, daß die experimentelle Variable einige Subjekte in der einen Richtung und andere Subjekte in der entgegengesetzten Richtung beeinflußt. In diesem Test wird auf extreme Antworten im Vergleich zu einer Kontrollgruppe geprüft. Dieser Test konzentriert sich auf die Spannweite der Kontrollgruppe und ist ein Maß dafür, wie stark die Spannweite durch die extremen Werte in der experimentellen Gruppe beeinflußt wird, wenn sie mit der Kontrollgruppe verbunden werden. Die Kontrollgruppe wird durch den Wert

der Gruppe 1 im Dialogfeld "Zwei unabhängige Stichproben: Gruppen definieren" bestimmt. Die Beobachtungen aus beiden Gruppen werden kombiniert und einem Rang zugeordnet. Die Spanne der Kontrollgruppe wird als die Differenz zwischen den Rängen der größten und kleinsten Werte in der Kontrollgruppe plus 1 berechnet. Da zufällige Ausreißer den Bereich der Spannweite leicht verzerren können, werden 5 % der Kontrollfälle automatisch an jedem Ende weggelassen.

Zwei unabhängige Stichproben: Gruppen definieren

Abbildung 35-15
Dialogfeld "Zwei unabhängige Stichproben: Gruppen definieren"



Um die Datei in zwei Gruppen oder Stichproben aufzuteilen, geben Sie eine ganze Zahl für Gruppe 1 und eine weitere Zahl für Gruppe 2 ein. Fälle mit anderen Werten werden aus der Analyse ausgeschlossen.

Zwei unabhängige Stichproben: Optionen

Abbildung 35-16
Dialogfeld "Zwei unabhängige Stichproben: Optionen"

Statistik. Sie können eine oder beide Auswertungsstatistiken wählen.

- **Deskriptive Statistik.** Zeigt Mittelwert, Standardabweichung, Minimum, Maximum und Anzahl der nichtfehlenden Fälle an.
- **Quartile.** Zeigt die Werte an, die den 25., 50. und 75. Perzentilen entsprechen.

Fehlende Werte. Bestimmt die Verarbeitung fehlender Werte.

- **Fallausschluß Test für Test.** Werden mehrere Tests festgelegt, so wird jeder Test einzeln auf fehlende Werte geprüft.
- **Listenweiser Fallausschluß.** Fälle mit fehlenden Werten für eine Variable werden aus allen Analysen ausgeschlossen.

Zusätzliche Funktionen beim Befehl NPAR TESTS (Tests bei zwei unabhängigen Stichproben)

Mit dem Unterbefehl `MOSES` der SPSS-Befehlssprache kann die Anzahl der Fälle angegeben werden, die für den Moses-Test getrimmt werden sollen.

Vollständige Informationen zur Syntax finden Sie in der *SPSS Command Syntax Reference*.

Tests bei zwei verbundenen Stichproben

Die Prozedur "Tests bei zwei verbundenen Stichproben" vergleicht die Verteilungen von zwei Variablen.

Beispiel. Erhalten Familien, die ihr Haus verkaufen, im allgemeinen den geforderten Preis? Wenn Sie den Wilcoxon-Test auf die Daten von 10 Häusern anwenden, könnten Sie beispielsweise feststellen, daß sieben Familien weniger als den geforderten Preis, eine Familie mehr als den geforderten Preis und zwei Familien den geforderten Preis erhielten.

Statistiken. Mittelwert, Standardabweichung, Minimum, Maximum, Anzahl der nichtfehlenden Fälle und Quartile. Tests: Wilcoxon-Test, Vorzeichentest, McNemar.

Daten. Verwenden Sie numerische Variablen, die geordnet werden können.

Annahmen. Obwohl keine bestimmten Verteilungen für die beiden Variablen vorausgesetzt werden, wird die Verteilung der Grundgesamtheit der gepaarten Differenzen als symmetrisch angenommen.

Abbildung 35-17
Ausgabe für "Zwei verbundene Stichproben"

Ränge

		N	Mittlerer Rang	Rangsumme
Verkaufspreis - Angebotspreis	Negative Ränge	7[1]	4.93	34.50
	Positive Ränge	1[2]	1.50	1.50
	Bindungen	2[3]		
	Gesamt	10		

1. Verkaufspreis < Angebotspreis
2. Verkaufspreis > Angebotspreis
3. Angebotspreis = Verkaufspreis

Statistik für Test[2]

	Verkaufspreis - Angebotspreis
Z	-2.313[1]
Asymptotische Signifikanz (2-seitig)	.021

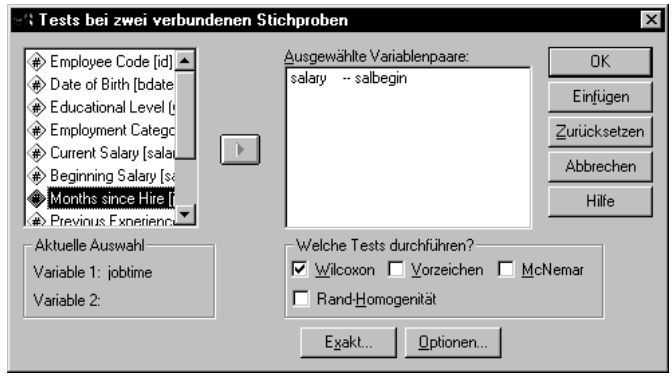
1. Basiert auf positiven Rängen.
2. Wilcoxon-Test

So lassen Sie Tests bei zwei verbundenen Stichproben berechnen

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Nichtparametrische Tests
    Zwei verbundene Stichproben…

Abbildung 35-18
Dialogfeld "Tests bei zwei verbundenen Stichproben"



- Wählen Sie mindestens ein Paar von Variablen aus. Gehen Sie dabei folgendermaßen vor:
  - Klicken Sie auf beide Variablen. Die erste Variable wird im Gruppenfeld "Aktuelle Auswahl" als *Variable 1*, die zweite Variable als *Variable 2* angezeigt.
  - Klicken Sie auf die Pfeilschaltfläche, um das Paar in die Liste "Ausgewählten Variablenpaare" zu verschieben. Sie können mehrere Variablenpaare auswählen. Um ein Variablenpaar aus der Analyse zu entfernen, wählen Sie es in der Liste "Ausgewählte Variablenpaare" aus, und klicken Sie auf die Pfeilschaltfläche.

Typen von Tests bei zwei verbundenen Stichproben

Die Tests in diesem Abschnitt vergleichen die Verteilungen von zwei verbundenen Variablen. Der geeignete Test hängt vom jeweiligen Datentyp ab.

Falls Ihre Daten stetig sind, verwenden Sie den Vorzeichentest oder den Wilcoxon-Test. Der **Vorzeichentest** berechnet für alle Fälle die Differenzen zwischen den beiden Variablen und klassifiziert sie als positiv, negativ oder verbunden. Falls die beiden Variablen ähnlich verteilt sind, unterscheidet sich die Zahl der positiven und negativen Differenzen nicht signifikant. Der **Wilcoxon-Test** berücksichtigt sowohl Informationen über Vorzeichen der Differenzen als auch die Größe der Differenzen zwischen den Paaren. Da der Wilcoxon-Test mehr Informationen über die Daten aufnimmt, kann er mehr leisten als der Vorzeichentest.

Falls Sie mit binären Daten arbeiten, verwenden Sie den **McNemar-Test**. Dieser Test wird üblicherweise bei Meßwiederholungen verwendet, wenn jede Antwort eines Subjektes doppelt abgerufen wird, einmal bevor ein festgelegtes Ereignis eintritt und einmal danach. Der McNemar-Test bestimmt, ob die Antwortrate am Anfang (vor dem Ereignis) gleich der Antwortrate am Ende (nach dem Ereignis) ist. Dieser Test ist für das Erkennen von Änderungen bei Antworten nützlich, die durch experimentelle Einflußnahme in sogenannten "Vorher-und-nachher-Designs" entstanden sind.

Falls Sie mit kategorialen Daten arbeiten, verwenden Sie den **Rand-Homogenitätstest**. Dieser Test ist eine Erweiterung des McNemar-Tests von binären Variablen auf multinomiale Variablen. Mithilfe dieses Tests wird unter Verwendung der Chi-Quadrat-Verteilung überprüft, ob Änderungen bei den Antworten vorliegen. Dies ist nützlich, um zu ermitteln, ob die Änderungen in sogenannten "Vorher-und-nachher-Designs" durch experimentelle Einflußnahme verursacht werden. Der Rand-Homogenitätstest ist nur verfügbar, wenn Sie die SPSS Exact Tests installiert haben.

Optionen für Tests bei zwei verbundenen Stichproben

Abbildung 35-19
Dialogfeld "Zwei verbundene Stichproben: Optionen"

Statistik. Sie können eine oder beide Auswertungsstatistiken wählen.

- **Deskriptive Statistik.** Zeigt Mittelwert, Standardabweichung, Minimum, Maximum und Anzahl der nichtfehlenden Fälle an.
- **Quartile.** Zeigt die Werte an, die den 25., 50. und 75. Perzentilen entsprechen.

Fehlende Werte. Bestimmt die Verarbeitung fehlender Werte.

- **Fallausschluß Test für Test.** Werden mehrere Tests festgelegt, so wird jeder Test einzeln auf fehlende Werte geprüft.
- **Listenweiser Fallausschluß.** Fälle mit fehlenden Werten für eine Variable werden aus allen Analysen ausgeschlossen.

Zusätzliche Funktionen beim Befehl NPAR TESTS (zwei verbundene Stichproben)

Mit der SPSS-Befehlssprache können Sie außerdem eine Variable mit jeder Variable auf einer Liste überprüfen.

Vollständige Informationen zur Syntax finden Sie in der *SPSS Command Syntax Reference*.

Tests bei mehreren unabhängigen Stichproben

Mit der Prozedur "Tests bei mehreren unabhängigen Stichproben" werden zwei oder mehrere Fallgruppen einer Variablen verglichen.

Beispiel. Unterscheiden sich 100-Watt-Glühlampen dreier Marken in ihrer durchschnittlichen Lebensdauer? Mit der einfaktoriellen Varianzanalyse nach Kruskal-Wallis könnten Sie feststellen, daß die drei Marken sich in ihrer durchschnittlichen Lebensdauer unterscheiden.

Statistiken. Mittelwert, Standardabweichung, Minimum, Maximum, Anzahl der nichtfehlenden Fälle und Quartile. Tests: Kruskal-Wallis-*H*, Median.

Daten. Verwenden Sie numerische Variablen, die geordnet werden können.

Annahmen. Verwenden Sie unabhängige Zufallsstichproben. Für den Kruskal-Wallis-*H*-Test sind Stichproben erforderlich, die sich in ihrer Form ähneln.

Abbildung 35-20
Ausgabe von Tests bei mehreren unabhängigen Stichproben

Ränge

			N	Mittlerer Rang
Stunden	Marke	Marke A	10	15,20
		Marke B	10	25,50
		Marke C	10	5,80
		Gesamt	30	

Statistik für Test[2]

	Tragezeit in Tagen
Mann-Whitney-U	14.000
Wilcoxon-W	69.000
Z	-2.721
Asymptotische Signifikanz (2-seitig)	.007
Exakte Signifikanz [2*(1-seitig Sig.)]	.005[1]

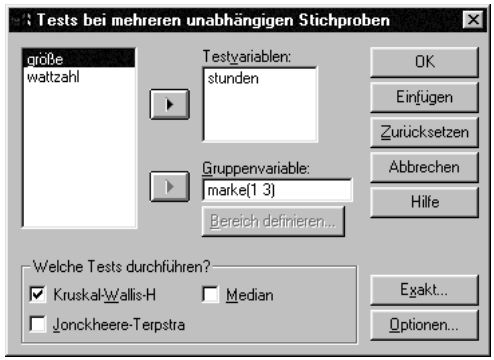
1. Nicht für Bindungen korrigiert.
2. Gruppenvariable: Zahnspangenmodell

So lassen Sie Tests für mehrere unabhängige Stichproben berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Nichtparametrische Tests
    K unabhängige Stichproben...

Abbildung 35-21
Dialogfeld "Tests bei mehreren unabhängigen Stichproben"



- Wählen Sie mindestens eine numerische Variable aus.
- Wählen Sie eine Gruppenvariable aus, und klicken Sie auf Bereich definieren, um die ganzzahligen Minimal- und Maximalwerte der Gruppenvariablen festzulegen.

Tests bei mehreren unabhängigen Stichproben: Welche Tests durchführen?

Sie können mit drei Tests bestimmen, ob mehrere unabhängige Stichproben aus derselben Grundgesamtheit stammen. Mit dem Kruskal-Wallis-*H*-Test, dem Mediantest und dem Jonckheere-Terpstra-Test können Sie prüfen, ob mehrere unabhängige Stichproben aus derselben Grundgesamtheit stammen.

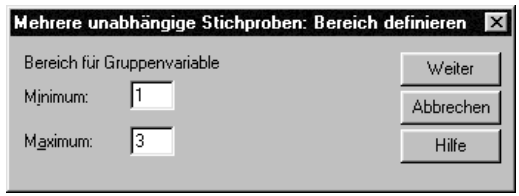
Der **Kruskal-Wallis-H-Test**, eine Erweiterung des Mann-Whitney-*U*-Tests, ist die nichtparametrische Entsprechung der einfaktoriellen Varianzanalyse und erkennt Unterschiede in der Lage der Verteilung. Der **Mediantest**, der allgemeiner, aber nicht so leistungsstark ist, erkennt Unterschiede von Verteilungen in Lage und Form. Der Kruskal-Wallis-*H*-Test und der Mediantest setzen voraus, dass keine *a-priori*-Ordnung der *k* Grundgesamtheiten vorliegt, aus denen die Stichproben gezogen werden.

Wenn eine natürliche *a-priori*-Ordnung (aufsteigend oder absteigend) der *k* Grundgesamtheiten *besteht*, ist der **Jonckheere-Terpstra-Test** leistungsfähiger. Die *k* Grundgesamtheiten könnten zum Beispiel *k* ansteigende Temperaturen darstellen. Die Hypothese, daß unterschiedliche Temperaturen die gleiche Verteilung von Antworten erzeugen, wird gegen die Alternative getestet, daß mit Zunahme der Temperatur die Größe der Antwort zunimmt. Hierbei ist die alternative Hypothese geordnet, deshalb ist der Jonckheere-Terpstra-Test für diesen Test am besten geeignet.

Der Jonckheere-Terpstra-Test ist nur verfügbar, wenn Sie SPSS Exact Tests installiert haben.

Tests bei mehreren unabhängigen Stichproben: Bereich definieren

Abbildung 35-22
Dialogfeld "Mehrere unabhängige Stichproben: Bereich definieren"



Um den Bereich zu definieren, geben Sie für Minimum und Maximum ganzzahlige Werte ein, die der niedrigsten und höchsten Kategorie der Gruppenvariablen entsprechen. Der Minimalwert muß kleiner sein als der Maximalwert. Wenn Sie zum Beispiel als Minimum 1 und als Maximum 3 angeben, werden nur die ganzzahligen Werte von 1 bis 3 verwendet. Das Minimum muß kleiner als das Maximum sein. Beide Werte müssen angegeben werden.

Tests bei mehreren unabhängigen Stichproben: Optionen

Abbildung 35-23
Dialogfeld "Mehrere unabhängige Stichproben: Optionen"

Statistik. Sie können eine oder beide Auswertungsstatistiken wählen.

- **Deskriptive Statistik.** Zeigt Mittelwert, Standardabweichung, Minimum, Maximum und Anzahl der nichtfehlenden Fälle an.
- **Quartile.** Zeigt die Werte an, die den 25., 50. und 75. Perzentilen entsprechen.

Fehlende Werte. Bestimmt die Verarbeitung fehlender Werte.

- **Fallausschluß Test für Test.** Werden mehrere Tests festgelegt, so wird jeder Test einzeln auf fehlende Werte geprüft.
- **Listenweiser Fallausschluß.** Fälle mit fehlenden Werten für eine Variable werden aus allen Analysen ausgeschlossen.

Zusätzliche Funktionen beim Befehl NPAR TESTS (K unabhängige Stichproben)

In der SPSS-Befehlssprache haben Sie außerdem die Möglichkeit, mit dem Unterbefehl `MEDIAN` einen anderen Wert als den beobachteten Median für den Mediantest festzulegen.

Vollständige Informationen zur Syntax finden Sie in der *SPSS Command Syntax Reference*.

Tests bei mehreren verbundenen Stichproben

Bei der Prozedur "Tests bei mehreren verbundenen Stichproben" werden die Verteilungen von zwei oder mehr Variablen verglichen.

Beispiel. Genießen die Berufsgruppen Ärzte, Anwälte, Polizisten oder Lehrer in der Öffentlichkeit ein unterschiedliches Ansehen? Zehn Personen wurden gebeten, diese vier Berufsgruppen in der Reihenfolge ihres Ansehens anzuordnen. Der Test nach Friedman zeigt, dass diese vier Berufsgruppen in der Öffentlichkeit tatsächlich ein unterschiedliches Ansehen genießen.

Statistiken. Mittelwert, Standardabweichung, Minimum, Maximum, Anzahl der nichtfehlenden Fälle und Quartile. Tests: Friedman, Kendall-*W* und Cochran-*Q*.

Daten. Verwenden Sie numerische Variablen, die geordnet werden können.

Annahmen. Nichtparametrische Tests erfordern keine Annahmen über die Form der zugrundeliegenden Verteilung. Verwenden Sie abhängige Zufallsstichproben.

Abbildung 35-24
Ausgabe für Tests bei mehreren verbundenen Stichproben

Ränge

	Mittlerer Rang
Arzt	1.50
Anwalt	2.50
Polizist	3.40
Lehrer	2.60

Statistik für Test[2]

	Verkaufspreis - Angebotspreis
Z	-2.313[1]
Asymptotische Signifikanz (2-seitig)	.021

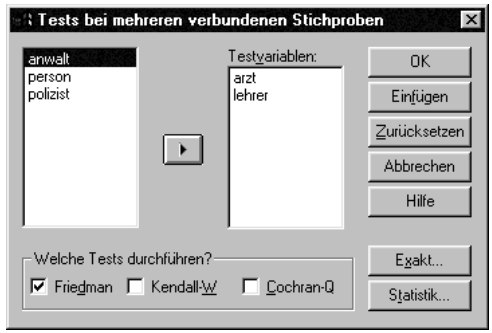
1. Basiert auf positiven Rängen.
2. Wilcoxon-Test

So lassen Sie Tests bei mehreren verbundenen Stichproben berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
Nichtparametrische Tests
K verbundene Stichproben...

Abbildung 35-25
Dialogfeld "Tests bei mehreren verbundenen Stichproben"



▶ Wählen Sie zwei oder mehr numerische Testvariablen aus.

Tests bei mehreren verbundenen Stichproben: Welche Tests durchführen?

Sie können die Verteilung von verschiedenen verbundenen Variablen mit drei Tests vergleichen.

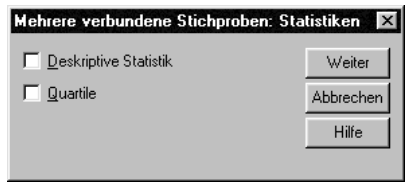
Der **Test** nach **Friedman** stellt das nichtparametrische Äquivalent eines Designs mit Meßwiederholungen bei einer Stichprobe bzw. eine Zweifach-Varianzanalyse mit einer Beobachtung pro Zelle dar. Der Friedman-Test überprüft die Nullhypothese, wonach die *k* verbundenen Variablen aus derselben Grundgesamtheit stammen. Für jeden Fall werden den *k* Variablen Rangzahlen von 1 bis *k* zugewiesen. Die Teststatistik wird auf der Grundlage dieser Ränge durchgeführt.

Das **Kendall-W** stellt eine Normalisierung der Statistik nach Friedman dar. Das Kendall-*W* kann als Konkordanzkoeffizient interpretiert werden, der ein Maß für die Übereinstimmung der Prüfer darstellt. Jeder Fall ist ein Richter oder Prüfer, und jede Variable ist ein zu beurteilendes Objekt oder eine zu beurteilende Person. Die Rangsumme jeder Variablen wird berechnet. Das Kendall-*W* liegt im Bereich von 0 (keine Übereinstimmung) bis 1 (vollständige Übereinstimmung).

Das **Cochran-Q** entspricht vollständig dem Friedman-Test. Es wird jedoch angewendet, wenn alle Antworten binär sind. Dieser Test stellt eine Erweiterung des McNemar-Tests auf *k* Stichproben dar. Das Cochran-*Q* überprüft die Hypothese, daß mehrere verbundene dichotome Variablen denselben Mittelwert aufweisen. Die Variablenwerte beziehen sich auf dasselbe Individuum oder auf zusammengehörige Individuen.

Tests bei mehreren verbundenen Stichproben: Statistiken

Abbildung 35-26
Dialogfeld "Mehrere verbundene Stichproben: Statistiken"



Sie können Statistiken auswählen.

- **Deskriptive Statistik.** Zeigt Mittelwert, Standardabweichung, Minimum, Maximum und Anzahl der nichtfehlenden Fälle an.
- **Quartile.** Zeigt die Werte an, die den 25., 50. und 75. Perzentilen entsprechen.

Zusätzliche Funktionen beim Befehl NPAR TESTS (K verbundene Stichproben)

Vollständige Informationen zur Syntax finden Sie im *SPSS Syntax Reference Guide*.

Kapitel
36
Analyse von Mehrfachantworten

Sie können für die Analyse von Sets aus dichotomen Variablen und von Sets aus kategorialen Variablen zwei Prozeduren verwenden. Mit der Prozedur "Mehrfachantworten: Häufigkeiten" können Sie Häufigkeitstabellen erstellen. Mit der Prozedur "Mehrfachantworten: Kreuztabellen" werden zwei- oder dreidimensionale Kreuztabellen angezeigt. Sie müssen Mehrfachantworten-Sets definieren, ehe Sie mit einer der Prozeduren beginnen.

Beispiel. Dieses Beispiel veranschaulicht den Gebrauch von Mehrfachantworten in einer Marktforschungsanalyse. Die hier verwendeten Daten sind frei erfunden und dürfen nicht als real interpretiert werden. Eine Fluggesellschaft führt eine Umfrage unter den Passagieren einer bestimmten Flugroute durch, um Informationen über konkurrierende Fluggesellschaften zu erhalten. In diesem Beispiel möchte American Airlines in Erfahrung bringen, welche anderen Fluggesellschaften ihre Passagiere auf der Route Chicago-New York nutzen und welche Rolle der Flugplan sowie der Service bei der Auswahl der Fluggesellschaft spielen. Der Flugbegleiter händigt jedem Passagier beim Einsteigen in die Maschine einen kurzen Fragebogen aus. Die erste Frage lautet: "Kreuzen Sie bitte alle der folgenden Fluggesellschaften an, mit denen Sie diese Route in den letzten sechs Monaten geflogen sind: American, United, TWA, USAir und andere." Dies ist eine Frage, die mit Mehrfachantworten beantwortet werden kann, weil jeder Passagier mehr als eine Antwort ankreuzen kann. Sie können diese Frage aber nicht direkt kodieren, weil eine SPSS-Variable für jeden Fall nur einen Wert annehmen kann. Sie müssen mehrere Variablen verwenden, um die Antworten zu jeder Frage zu erfassen. Dazu haben Sie zwei Möglichkeiten. Eine Möglichkeit besteht darin, zu jeder Antwortmöglichkeit eine entsprechende Variable zu definieren, also zum Beispiel "American", "United", "TWA", "USAir" und "andere". Wenn ein Passagier "United" ankreuzt, wird der Variablen *united* der Code 1 zugewiesen, sonst erhält diese den Code 0. Bei dieser Methode werden Variablen in **mehreren Dichotomien** erfaßt. Eine andere Möglichkeit stellt das Erfassen der Antworten in **mehreren Kategorien** dar, bei der Sie die maximale

Anzahl möglicher Antworten auf die Frage schätzen und eine entsprechende Anzahl von Variablen festlegen. Hierbei wird die verwendete Fluggesellschaft mit Hilfe eines Codes angegeben. Beim Durchsehen einer Stichprobe von Fragebögen stellen Sie vielleicht fest, daß in den letzten sechs Monaten kein Passagier mit mehr als drei verschiedenen Fluggesellschaften auf dieser Route geflogen ist. Außerdem bemerken Sie, daß aufgrund der Liberalisierung des Luftverkehrs 10 weitere Fluggesellschaften in der Kategorie "Andere" genannt sind. Mit der Methode für mehrere Kategorien würden Sie drei Variablen definieren. Jede würde wie folgt kodiert sein: 1 = *american*, 2 = *united*, 3 = *twa*, 4 = *usair*, 5 = *delta* usw. Wenn ein Passagier "American" und "TWA" ankreuzt, wird der ersten Variablen der Code 1 zugewiesen, der zweiten der Code 3 und der dritten ein Code für fehlende Werte. Ein anderer Passagier hat vielleicht "American" und "Delta" angekreuzt. Dementsprechend wird der ersten Variablen der Code 1, der zweiten der Code 5 und der dritten ein Code für fehlende Werte zugewiesen. Dagegen führt die Methode für mehrfache Dichotomie zu 14 verschiedenen Variablen. Obwohl beide Methoden für dieses Umfragebeispiel geeignet sind, hängt die Wahl der Methode von der Verteilung der Antworten ab.

Mehrfachantworten: Sets definieren

Mit der Prozedur "Mehrfachantworten: Sets definieren" können Sie elementare Variablen in Sets aus dichotomen Variablen und Sets aus kategorialen Variablen gruppieren. Für diese Sets können Sie Häufigkeitstabellen und Kreuztabellen erstellen. Sie können bis zu 20 Mehrfachantworten-Sets definieren. Jedes Set muß über einen eigenen eindeutigen Namen verfügen. Sie können ein Set entfernen, indem Sie es in der Liste der Mehrfachantworten-Sets markieren und anschließend auf Entfernen klicken. Sie können ein Set ändern, indem Sie es in der Liste markieren, die Charakteristiken der Set-Definition ändern und anschließend auf Ändern klicken.

Sie können die elementaren Variablen als Dichotomien oder als Kategorien definieren. Wenn Sie dichotome Variablen verwenden möchten, aktivieren Sie das Optionsfeld Dichotomien, um ein Set aus dichotomen Variablen zu erstellen. Geben Sie für "Gezählter Wert" eine ganze Zahl ein. Jede Variable, bei welcher der gezählte Wert mindestens einmal auftritt, wird zu einer Kategorie des Sets aus dichotomen Variablen. Aktivieren Sie das Optionsfeld Kategorien, um ein Set aus kategorialen Variablen zu erstellen, das den gleichen Wertebereich wie die Komponentenvariablen umfaßt. Geben Sie ganzzahlige Werte für die Minimal- und Maximalwerte des Bereichs für die Kategorien des Sets aus kategorialen Variablen ein. SPSS

bildet die Summe aller unterschiedlichen ganzzahligen Werte im Bereich aller Komponentenvariablen. Leere Kategorien werden nicht in Tabellen übernommen.

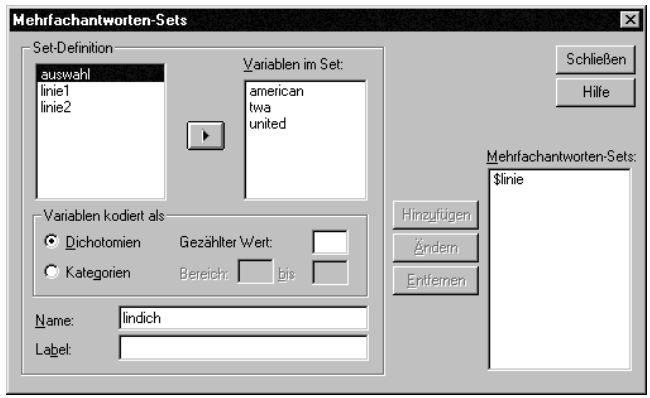
Sie müssen jedem Mehrfachantworten-Set einen eindeutigen Namen zuweisen, der aus bis zu sieben Zeichen bestehen darf. SPSS stellt dem von Ihnen zugewiesenen Namen das Dollarzeichen ($) als Präfix voran. Die folgenden reservierten Namen dürfen Sie nicht verwenden: *casenum*, *sysmis*, *jdate*, *date*, *time*, *length* und *width*. Der Name des Mehrfachantworten-Sets ist nur zur Verwendung in Mehrfachantworten-Prozeduren vorgesehen. In anderen Prozeduren können Sie sich nicht auf Namen von Mehrfachantworten-Sets beziehen. Wahlweise können Sie für das Mehrfachantworten-Set ein aussagekräftiges Variablenlabel eingeben. Das Label kann bis zu 40 Zeichen lang sein.

So definieren Sie Mehrfachantworten-Sets

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Mehrfachantworten
    Sets definieren…

Abbildung 36-1
Dialogfeld "Mehrfachantworten-Sets"



▶ Wählen Sie mindestens zwei Variablen aus.

▶ Wenn Ihre Variablen als Dichotomien kodiert sind, geben Sie an, welcher Wert gezählt werden soll. Wenn Ihre Variablen als Kategorien kodiert sind, legen Sie den Bereich für die Kategorien fest.

- Geben Sie einen eindeutigen Namen für jedes Mehrfachantworten-Set ein.
- Klicken Sie auf Hinzufügen, um das Mehrfachantworten-Set zur Liste der definierten Sets hinzuzufügen.

Mehrfachantworten: Häufigkeiten

Mit der Prozedur "Mehrfachantworten: Häufigkeiten" erstellen Sie Häufigkeitstabellen für Mehrfachantworten-Sets. Zuvor müssen Sie mindestens ein Mehrfachantworten-Set definieren (siehe "Mehrfachantworten: Sets definieren").

Bei Sets aus dichotomen Variablen entsprechen die in der Ausgabe gezeigten Kategorienamen den Variablenlabels, die für die elementaren Variablen in der Gruppe festgelegt wurden. Wenn keine Variablenlabels festgelegt wurden, werden die Variablennamen als Labels verwendet. Bei Sets aus kategorialen Variablen entsprechen die Kategoriebeschriftungen den Wertelabels der ersten Variable in der Gruppe. Wenn Kategorien, die bei der ersten Variable fehlen, bei anderen Variablen in der Gruppe vorhanden sind, müssen Sie ein Wertelabel für die fehlenden Kategorien festlegen.

Fehlende Werte. Fälle mit fehlenden Werten werden jeweils für einzelne Tabellen ausgeschlossen. Sie können aber auch eine oder beide der folgenden Möglichkeiten auswählen:

- **Für dichotome Variablen Fälle listenweise ausschließen.** Fälle, bei denen Werte einer beliebigen Variablen fehlen, werden aus der Tabelle des Sets aus dichotomen Variablen ausgeschlossen. Dies gilt nur für Mehrfachantworten-Sets, die als Sets aus dichotomen Variablen definiert wurden. In der Standardeinstellung gilt ein Fall in einem Set aus dichotomen Variablen als fehlend, wenn keine der Variablen des Falls den gezählten Wert enthält. Fälle mit fehlenden Werten für nur einige, aber nicht alle der Variablen werden in die Tabellen der Gruppe aufgenommen, wenn mindestens eine Variable den gezählten Wert enthält.
- **Für kategoriale Variablen Fälle listenweise ausschließen.** Fälle, bei denen Werte einer beliebigen Variablen fehlen, werden aus der Tabelle des Sets aus kategorialen Variablen ausgeschlossen. Dies gilt nur für Mehrfachantworten-Sets, die als Sets aus kategorialen Variablen definiert wurden. In der Standardeinstellung gilt ein Fall in einem Set aus kategorialen Variablen nur als fehlend, wenn keine der Komponenten des Falls gültige Werte innerhalb des definierten Bereichs enthält.

Beispiel. Jede SPSS-Variable, die aus einer Frage in einer Umfrage erstellt wurde, stellt eine elementare Variable dar. Zum Analysieren der Mehrfachantworten müssen Sie die Variablen in einem der beiden möglichen Typen von Mehrfachantworten-Sets zusammenfassen: in einem Set aus dichotomen Variablen oder in einem Set aus kategorialen Variablen. Wenn zum Beispiel in einer Umfrage ermittelt wurde, mit welcher von drei verschiedenen Fluggesellschaften (American, United und TWA) die befragten Personen in den letzten sechs Monaten geflogen sind, und Sie haben dichotome Variablen verwendet und ein **Set aus dichotomen Variablen** definiert, dann würde jede der drei Variablen im Set zu einer Kategorie der Gruppenvariablen werden. Die Angaben zu Anzahl und Prozentwert für jede Fluggesellschaft werden zusammen in einer Häufigkeitstabelle angezeigt. Wenn Sie feststellen, daß keiner der Befragten mit mehr als zwei Fluggesellschaften geantwortet hat, können Sie zwei Variablen erstellen, die jeweils einen von drei Codes annehmen können. Dabei stellt jeder Code eine Fluggesellschaft dar. Wenn Sie ein **Set aus kategorialen Variablen** definieren, stellen die Werte in der Tabelle die Anzahl von gleichen Codes in den elementaren Variablen dar. Das resultierende Set von Werten entspricht denen für jede einzelne der elementaren Variablen. So entsprechen beispielsweise 30 Antworten mit “United” der Summe von fünf Antworten mit “United” für “Fluglinie 1” und 25 Antworten mit “United” für “Fluglinie 2”. Die Angaben zu Anzahl und Prozentwert für jede Fluggesellschaft werden zusammen in einer Häufigkeitstabelle angezeigt.

Statistiken. Häufigkeitstabellen mit den Häufigkeiten, Prozentsätzen der Antworten, Prozentsätzen der Fälle, der Anzahl gültiger Fälle und der Anzahl fehlender Fälle.

Daten. Verwenden Sie Mehrfachantworten-Sets.

Annahmen. Die Häufigkeiten und Prozentsätze geben nützliche Beschreibungen für Daten mit beliebigen Verteilungen.

Verwandte Prozeduren. Mit der Prozedur “Mehrfachantworten: Sets definieren” können Sie Mehrfachantworten-Sets definieren.

Abbildung 36-2
Ausgabe für "Mehrfachantworten: Häufigkeiten"

		Antworten		Prozentsatz der Fälle
		N	Prozent	
Name[a]	American	75	67,6%	92,6%
	TWA	6	5,4%	7,4%
	United	30	27,0%	37,0%
Gesamt		111	100,0%	137,0%

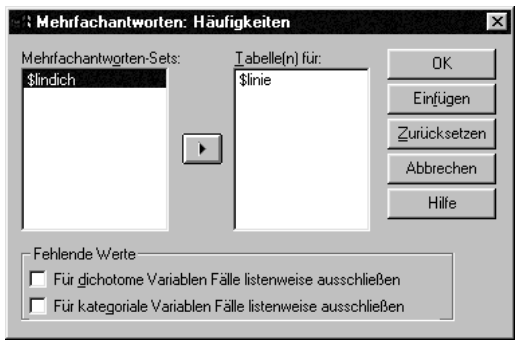
a. Dichotomiegruppe bei Wert 1 tabelliert.

So berechnen Sie Häufigkeiten mit Mehrfachantworten

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
Mehrfachantworten
Häufigkeiten...

Abbildung 36-3
Dialogfeld "Mehrfachantworten: Häufigkeiten"



▶ Wählen Sie mindestens ein Mehrfachantworten-Set aus.

Mehrfachantworten: Kreuztabellen

Mit der Prozedur "Mehrfachantworten: Kreuztabellen" können Kreuztabellen für definierte Mehrfachantworten-Sets, elementare Variablen oder eine Kombination dieser Elemente berechnet werden. Sie können außerdem Prozentwerte für Zellen basierend auf Fällen oder Antworten berechnen lassen, die Verarbeitung von fehlenden Werten ändern oder gepaarte Kreuztabellen erstellen lassen. Zuvor müssen

Sie mindestens ein Mehrfachantworten-Set definieren (siehe "So definieren Sie Mehrfachantworten-Sets").

Bei Sets aus dichotomen Variablen entsprechen die in der Ausgabe gezeigten Kategorienamen den Variablenlabels, die für die elementaren Variablen in der Gruppe festgelegt wurden. Wenn keine Variablenlabels festgelegt wurden, werden die Variablennamen als Labels verwendet. Bei Sets aus kategorialen Variablen entsprechen die Kategoriebeschriftungen den Wertelabels der ersten Variable in der Gruppe. Wenn Kategorien, die bei der ersten Variable fehlen, bei anderen Variablen in der Gruppe vorhanden sind, müssen Sie ein Wertelabel für die fehlenden Kategorien festlegen. In SPSS werden die Kategorienbeschriftungen für Spalten auf drei Zeilen mit bis zu acht Zeichen pro Zeile angezeigt. Wenn Sie vermeiden möchten, daß Wörter getrennt werden, können Sie die Anordnung von Zeilen und Spalten umdrehen oder die Labels neu festlegen.

Beispiel. Sowohl Sets aus dichotomen Variablen als auch Sets aus kategorialen Variablen können bei dieser Prozedur mit anderen Variablen in eine Kreuztabelle eingehen. Bei einer Befragung von Passagieren einer Fluglinie werden die Reisenden um die folgenden Informationen gebeten: Kreuzen Sie bitte alle der folgenden Fluggesellschaften an, mit denen Sie in den letzten sechs Monaten geflogen sind (American, United und TWA). Was ist wichtiger, wenn Sie einen Flug buchen: der Flugplan oder der Service? Wählen Sie nur eine Möglichkeit aus. Nachdem Sie die Daten als Dichotomien oder multiple Kategorien eingegeben und diese in einem Set zusammengefaßt haben, können Sie die Auswahl der Fluggesellschaften zusammen mit der Frage nach Service bzw. Flugplan als Kreuztabelle berechnen lassen.

Statistiken. Kreuztabellen mit Häufigkeiten pro Zelle, Zeile, Spalte und Gesamt sowie Prozentsätzen für Zellen, Zeilen, Spalten und Gesamt. Die Prozentwerte für die Zellen können auf Fällen oder auf Antworten basieren.

Daten. Verwenden Sie Mehrfachantworten-Sets oder numerische kategoriale Variablen.

Annahmen. Die Häufigkeiten und Prozentsätze geben nützliche Beschreibungen für Daten mit beliebigen Verteilungen.

Verwandte Prozeduren. Mit der Prozedur "Mehrfachantworten: Sets definieren" können Sie Mehrfachantworten-Sets definieren.

Abbildung 36-4
Ausgabe für "Mehrfachantworten: Kreuztabellen"

			Select airline because of		
			Schedule	Service	Gesamt
Name	American	Anzahl	41	34	75
	TWA	Anzahl	3	3	6
	United	Anzahl	27	3	30
Gesamt		Anzahl	44	37	81

Prozentsätze und Gesamtwerte beruhen auf den Befragten.
a. Dichotomiegruppe bei Wert 1 tabelliert.

So berechnen Sie Kreuztabellen mit Mehrfachantworten

- Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
Mehrfachantworten
Kreuztabellen...

Abbildung 36-5
Dialogfeld "Mehrfachantworten: Kreuztabellen"

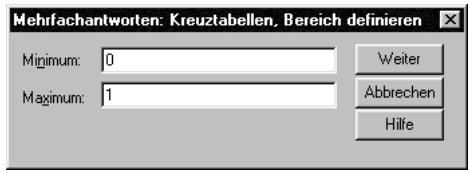
- Wählen Sie mindestens eine numerische Variable oder mindestens ein Mehrfachantworten-Set für jede Dimension der Kreuztabelle aus.

- Definieren Sie den Bereich jeder elementaren Variablen.

Außerdem können Sie eine Zweifach-Kreuztabelle für jede Kategorie einer Kontroll-Variablen oder eines Mehrfachantworten-Sets berechnen lassen. Wählen Sie mindestens einen Eintrag für die Liste “Schicht(en)” aus.

Mehrfachantworten: Kreuztabellen, Bereich definieren

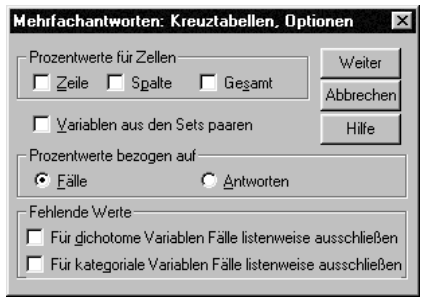
Abbildung 36-6
Dialogfeld “Mehrfachantworten: Kreuztabellen, Bereich definieren”



Für jede elementare Variable in der Kreuztabelle muß ein gültiger Wertebereich festgelegt werden. Geben Sie für die niedrigsten und höchsten Kategoriewerte, die in die Berechnung eingehen sollen, ganze Zahlen ein. Kategorien außerhalb des gültigen Bereichs werden aus der Analyse ausgeschlossen. Bei Werten innerhalb des einschließenden Bereichs wird von ganzen Zahlen ausgegangen, Stellen nach dem Komma werden abgeschnitten.

Mehrfachantworten: Kreuztabellen, Optionen

Abbildung 36-7
Dialogfeld “Mehrfachantworten: Kreuztabellen, Optionen”



Prozentwerte für Zellen. Die Zellenhäufigkeiten werden immer angezeigt. Sie können aber auch Spalten- und Zeilenprozentwerte sowie Prozentwerte für Zweifach-Tabellen (Gesamtwerte) anzeigen lassen.

Prozentwerte bezogen auf. Sie können festlegen, daß die Prozentsätze für die Zellen auf Fällen basieren. Diese Option ist nicht verfügbar, wenn Sie Variablen aus verschiedenen Sets von kategorialen Variablen paaren. Die Prozentsätze für die Zellen können außerdem auf den Antworten basieren. Bei Sets aus dichotomen Variablen entspricht die Anzahl der Antworten der Anzahl von gezählten Werten in allen Fällen. Bei Sets aus kategorialen Variablen entspricht die Anzahl der Antworten der Anzahl von Werten im festgelegten Bereich.

Fehlende Werte. Sie können eine oder beide der folgenden Möglichkeiten auswählen:

- **Für dichotome Variablen Fälle listenweise ausschließen.** Fälle, bei denen Werte einer beliebigen Variablen fehlen, werden aus der Tabelle des Sets aus dichotomen Variablen ausgeschlossen. Dies gilt nur für Mehrfachantworten-Sets, die als Sets aus dichotomen Variablen definiert wurden. In der Standardeinstellung gilt ein Fall in einem Set aus dichotomen Variablen als fehlend, wenn keine der Variablen des Falls den gezählten Wert enthält. Fälle mit fehlenden Werten für nur einige, aber nicht alle der Variablen werden in die Tabellen der Gruppe aufgenommen, wenn mindestens eine Variable den gezählten Wert enthält.
- **Für kategoriale Variablen Fälle listenweise ausschließen.** Fälle, bei denen Werte einer beliebigen Variablen fehlen, werden aus der Tabelle des Sets aus kategorialen Variablen ausgeschlossen. Dies gilt nur für Mehrfachantworten-Sets, die als Sets aus kategorialen Variablen definiert wurden. In der Standardeinstellung gilt ein Fall in einem Set aus kategorialen Variablen nur als fehlend, wenn keine der Komponenten des Falls gültige Werte innerhalb des definierten Bereichs enthält.

Die Standardeinstellung von SPSS sieht vor, daß beim Erstellen von Kreuztabellen für Sets aus kategorialen Variablen jede Variable in der ersten Gruppe mit jeder Variablen in der zweiten Gruppe gepaart wird und die Häufigkeiten für jede Zelle addiert werden. Deshalb können manche Antworten mehr als einmal in einer Tabelle vorkommen. Sie können die folgende Option auswählen:

Variablen aus den Sets paaren. Hiermit wird die erste Variable aus der ersten Gruppe mit der ersten Variable aus der zweiten Gruppe gepaart usw. Wenn Sie diese Option auswählen, basieren die relativen Häufigkeiten in den Zellen nicht auf den Fällen, sondern auf den Antworten. Bei Sets aus dichotomen Variablen und elementaren Variablen steht das Paaren nicht zur Verfügung.

Zusätzliche Funktionen beim Befehl MULT RESPONSE

Mit der SPSS-Befehlssprache verfügen Sie über die folgenden zusätzlichen Möglichkeiten:

- Mit dem Unterbefehl `BY` können Kreuztabellen mit bis zu fünf Dimensionen berechnet werden.
- Mit dem Unterbefehl `FORMAT` können die Optionen für die Ausgabeformatierung geändert werden. So können beispielsweise Wertelabels unterdrückt werden.

Vollständige Informationen zur Syntax finden Sie im *SPSS Syntax Reference Guide*.

Kapitel
37

Ergebnisberichte

Auflistungen von Fällen und deskriptive Statistiken sind wichtige Hilfsmittel zur Untersuchung und Darstellung von Daten. Mit dem Daten-Editor oder der Prozedur "Berichte" können Sie Fälle auflisten, mit den Prozeduren "Häufigkeiten" Häufigkeitszählungen und deskriptive Statistiken erstellen und mit der Prozedur "Mittelwert" Statistiken für Teilgrundgesamtheiten anfordern. In jeder dieser Prozeduren wird ein zur übersichtlichen Darstellung von Informationen geeignetes Format verwendet. Mit den Funktionen "Bericht in Zeilen" und "Bericht in Spalten" können Sie für Informationen auch ein anderes Format der Datendarstellung wählen.

Bericht in Zeilen

Mit der Funktion "Bericht in Zeilen" werden Berichte erstellt, in denen verschiedene Auswertungsstatistiken in Zeilen angegeben sind. Ebenso sind Listen von Fällen mit oder ohne Auswertungsstatistik verfügbar.

Beispiel. In einem Einzelhandelsunternehmen mit Filialen werden Informationen über Angestellte, Gehälter, Anstellungszeiten sowie Filiale und Abteilung jedes Beschäftigten in Datensätzen gespeichert. Sie können einen Bericht erstellen, der nach Filiale und Abteilung (Break-Variablen) aufgeteilte Informationen (Listen) zu den einzelnen Beschäftigten liefert und eine Auswertungsstatistik (zum Beispiel Durchschnittsgehalt) für jede Filiale, jedes Ressort und jede Abteilung einer Filiale enthält.

Datenspalten. Hier werden die Berichtsvariablen aufgelistet, für die Sie Fälle auflisten oder Auswertungsstatistiken erstellen möchten, und das Anzeigeformat der Datenspalten festgelegt.

Break-Spalten. Hier werden optionale Break-Variablen aufgelistet, die den Bericht in Gruppen aufteilen, und Einstellungen für die Auswertungsstatistik sowie Anzeigeformate für Break-Spalten festgelegt. Bei mehreren Break-Variablen wird für

jede Kategorie einer Break-Variablen eine getrennte Gruppe innerhalb der Kategorien der vorhergehenden Break-Variablen in der Liste erzeugt. Die Break-Variablen müssen diskrete kategoriale Variablen sein, welche die Fälle in eine begrenzte Anzahl von sinnvollen Kategorien aufteilen. Die Einzelwerte jeder Break-Variablen werden in einer getrennten Spalte links von allen Datenspalten angezeigt.

Bericht. Hiermit werden alle Merkmale eines Berichts festgelegt, einschließlich zusammenfassender Gesamtstatistiken, Anzeige der fehlenden Werte, Seitennumerierung und Titel.

Fälle anzeigen. Hiermit werden für jeden Fall die aktuellen Werte (oder Wertelabels) von den Variablen der Datenspalten angezeigt. Dadurch wird ein Listenbericht erzeugt, der wesentlich umfangreicher als ein Zusammenfassungsbericht sein kann.

Vorschau. Es wird nur die erste Seite des Berichtes angezeigt. Mit dieser Option erhalten Sie eine Vorschau auf das Format Ihres Berichts, ohne diesen komplett bearbeiten zu müssen.

Daten sind schon sortiert. Bei Berichten mit Break-Variablen muß die Datendatei vor dem Erstellen des Berichts nach den Werten der Break-Variablen sortiert werden. Wenn Ihre Datendatei bereits nach den Werten der Break-Variablen sortiert ist, können Sie durch Auswahl dieser Option Bearbeitungszeit einsparen. Diese Option ist besonders hilfreich, wenn Sie bereits einen Bericht für die Vorschau erstellt haben.

Beispielausgabe

Abbildung 37-1
Kombinierter Bericht mit Listen von Fällen und Auswertungsstatistiken

So erstellen Sie eine Zusammenfassung: Bericht in Zeilen

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Berichte
  Bericht in Zeilen...

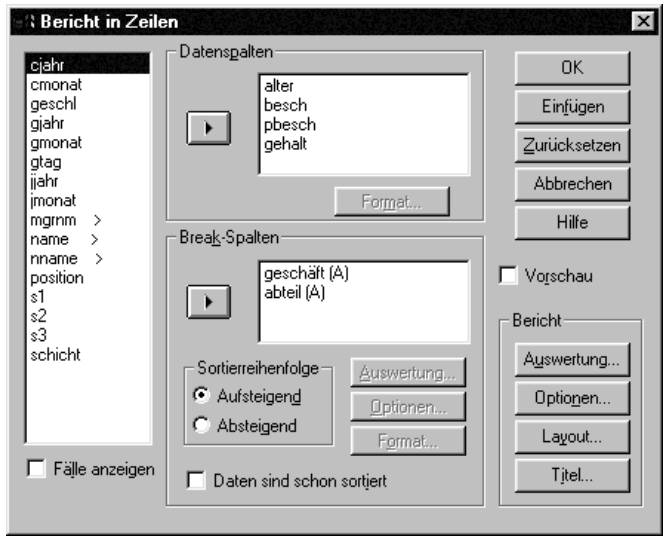
- Wählen Sie mindestens eine Variable für die Datenspalten aus. Für jede ausgewählte Variable wird eine Spalte im Bericht erzeugt.

- Wählen Sie bei sortierten und nach Untergruppen angezeigten Berichten mindestens eine Variable für die Break-Spalten aus.

- Bei Berichten mit Auswertungsstatistiken für Untergruppen, die durch Break-Variablen definiert wurden, wählen Sie in der Liste “Break-Spalten” die Break-Variablen aus, und klicken Sie im Gruppenfeld “Break-Spalten” auf Auswertung, um das (die) Auswertungsmaß(e) festzulegen.

- Bei Berichten mit zusammenfassenden Auswertungsstatistiken klicken Sie im Gruppenfeld “Bericht” auf Auswertung, um das (die) Auswertungsmaß(e) festzulegen.

Abbildung 37-2
Dialogfeld “Bericht in Zeilen”



Datenspaltenformat/Break-Format in Berichten

In den Format-Dialogfeldern werden Spaltentitel, Spaltenbreite, Textausrichtung sowie Anzeige der Datenwerte oder Wertelabels festgelegt. Mit “Datenspaltenformat” wird das Format der Datenspalten auf der rechten Seite des Berichtes festgelegt. Das Format der Break-Spalten auf der linken Seite wird mit “Break-Format” festgelegt.

Abbildung 37-3
Dialogfeld "Bericht: Datenspaltenformat"

Spaltentitel. Hiermit legen Sie den Spaltentitel für die ausgewählte Variable fest. Lange Titel werden in der Spalte automatisch umgebrochen. Verwenden Sie die Eingabetaste, um Zeilenumbrüche für Titel manuell einzufügen.

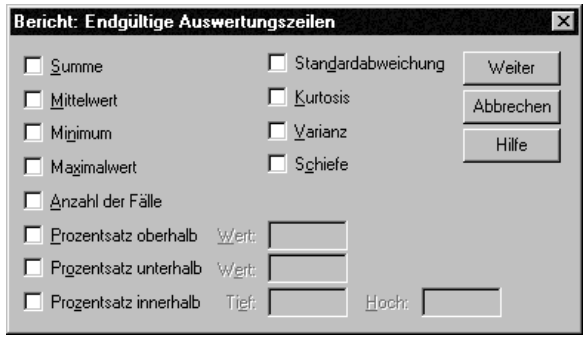
Position des Wertes in der Spalte. Hiermit wird für die ausgewählte Variable die Ausrichtung des Datenwerts oder Wertelabels in der Spalte festgelegt. Die Ausrichtung der Werte oder Labels hat keinen Einfluß auf die Ausrichtung der Spaltenüberschriften. Der Spalteninhalt kann entweder um eine festgelegte Anzahl von Zeichen eingerückt oder zentriert werden.

Spalteninhalt. Steuert die Anzeige von Datenwerten oder definierten Wertelabels der ausgewählten Variablen. Für Werte ohne definierte Wertelabels werden immer Datenwerte angezeigt. (Nicht verfügbar für Datenspalten in Bericht in Spalten.)

Bericht: Auswertungszeilen für/Endgültige Auswertungszeilen

Die beiden Dialogfelder für Auswertungszeilen legen Einstellungen für die Anzeige der Auswertungsstatistik für Break-Gruppen und für den gesamten Bericht fest. Mit "Auswertung" können Sie Einstellungen bezüglich der Untergruppenstatistik für jede durch die Break-Variablen definierte Kategorie vornehmen. Mit "Endgültige Auswertungszeilen" können Sie Einstellungen für die am Ende des Berichts angezeigte Gesamtstatistik vornehmen.

Abbildung 37-4
Dialogfeld "Bericht: Auswertung"

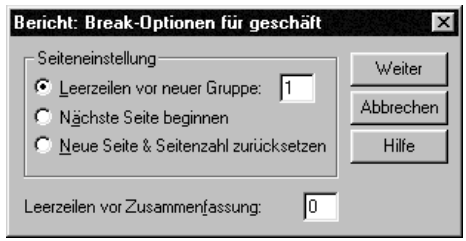


Die verfügbaren Auswertungsstatistiken sind Summe, Mittelwert, Minimum, Maximum, Anzahl der Fälle, Prozent der Fälle über oder unter einem festgelegten Wert, Prozent der Fälle innerhalb eines festgelegten Wertebereichs, Standardabweichung, Kurtosis, Varianz und Schiefe.

Bericht: Break-Optionen

Mit "Break-Optionen" werden Abstand und Seitenaufteilung der Informationen in den Break-Kategorien festgelegt.

Abbildung 37-5
Dialogfeld "Bericht: Break-Optionen"



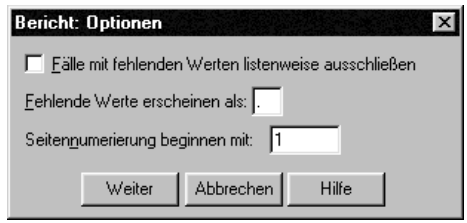
Seiteneinstellung. Hiermit werden Abstand und Seitenaufteilung für Kategorien der ausgewählten Break-Variablen festgelegt. Sie können eine Anzahl von Leerzeilen zwischen den Break-Kategorien festlegen oder eine Break-Kategorie an einen neuen Seitenanfang legen.

Leerzeilen vor Zusammenfassung. Hiermit legen Sie die Anzahl der Leerzeilen zwischen Beschriftungen oder Daten von Break-Kategorien und Auswertungsstatistiken fest. Dies bietet sich besonders für kombinierte Berichte mit Listen von einzelnen Fällen und Auswertungsstatistiken für Break-Kategorien an. In diesen Berichten können Sie Leerraum zwischen Listen von Fällen und Auswertungsstatistiken einfügen.

Bericht: Optionen

Mit "Bericht: Optionen" werden Behandlung und Anzeige der fehlenden Werte sowie Seitenaufteilung des Berichts festgelegt.

Abbildung 37-6
Dialogfeld "Bericht: Optionen"



Fälle mit fehlenden Werten listenweise ausschließen. Für jede der Berichtsvariablen werden sämtliche Fälle mit fehlenden Werten (im Bericht) ausgeschlossen.

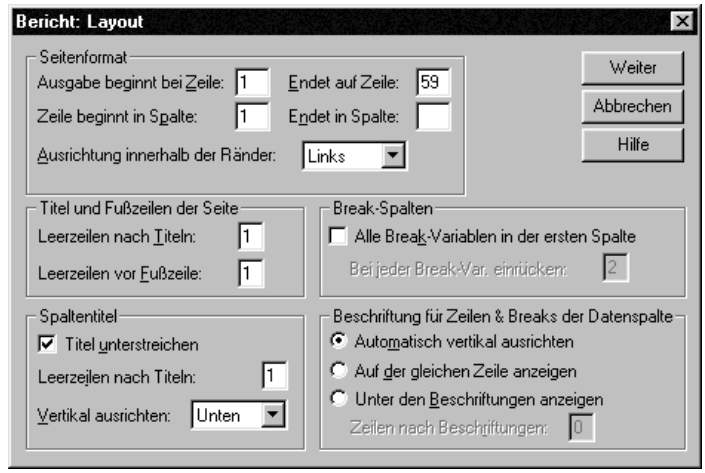
Fehlende Werte erscheinen als. Hier legen Sie das Symbol für fehlende Werte in der Datendatei fest. Das Symbol darf nur aus einem Zeichen bestehen und wird sowohl zur Darstellung **systemdefiniert fehlender** als auch **benutzerdefiniert fehlender** Werte verwendet.

Seitennumerierung beginnen mit. Mit dieser Option können Sie für die erste Seite des Berichts eine Seitennummer festlegen.

Bericht: Layout

Mit "Bericht: Layout" werden Breite und Länge jeder Berichtsseite, Seitenanordnung des Berichts sowie Einfügen von Leerzeilen und Beschriftungen festgelegt.

Abbildung 37-7
Dialogfeld "Bericht: Layout"



Seitenformat. Legt die Seitenränder, ausgedrückt in Zeilen (oben und unten) und Leerzeichen (links und rechts) sowie die Ausrichtung des Berichts innerhalb der Ränder fest.

Titel und Fußzeilen der Seite. Legt die Anzahl von Zeilen fest, welche die Kopf- und Fußzeile jeweils vom Text des Berichts trennen.

Break-Spalten. Hiermit wird die Anzeige der Break-Spalten festgelegt. Wenn mehrere Break-Variablen festgelegt wurden, können sie sich in getrennten Spalten oder in der ersten Spalte befinden. Das Anordnen aller Break-Variablen in der ersten Spalte erzeugt einen schmaleren Bericht.

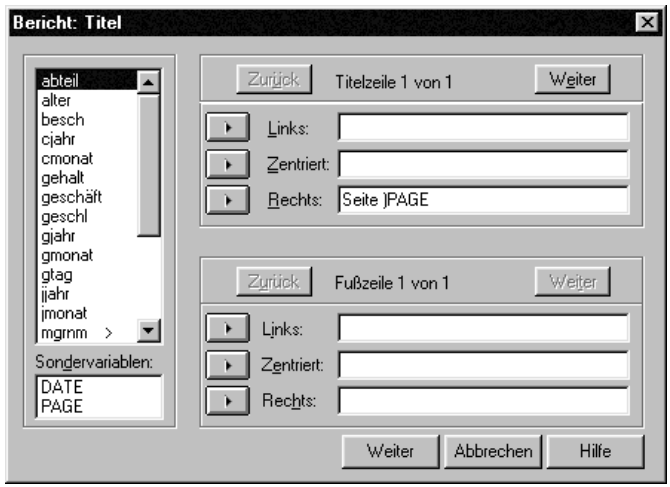
Spaltentitel. Legt die Anzeige von Spaltentiteln fest und umfaßt Unterstreichung des Titels, Anzahl von Leerzeilen zwischen Titel und Text des Berichts sowie die vertikale Ausrichtung.

Beschriftung für Zeilen & Breaks der Datenspalte. Steuert die Anordnung von Informationen in Datenspalten (Datenwerte und/oder Auswertungsstatistiken) bezüglich der Break-Beschriftungen zu Beginn jeder Break-Kategorie. Die erste Informationszeile in der Datenspalte kann entweder in der gleichen Zeile wie die Beschriftung der Break-Kategorie oder nach einer festgelegten Anzahl von Zeilen nach der Beschriftung der Break-Kategorie beginnen. (Nicht für Auswertungsberichte in Spalten verfügbar.)

Bericht: Titel

Im Dialogfeld “Bericht: Titel” werden Inhalt und Anordnung der Titel- und Fußzeilen des Berichts festgelegt. Sie können jeweils bis zu zehn Titel- und Fußzeilen festlegen, wobei in jeder Zeile linksbündige, zentrierte oder rechtsbündige Komponenten enthalten sein können.

Abbildung 37-8
Dialogfeld “Bericht: Titel”



Wenn Sie in Titeln oder Fußzeilen Variablen eingeben, wird das aktuelle Wertelabel oder der Wert der Variablen im Titel oder in der Fußzeile angezeigt. In Titeln wird das Wertelabel angezeigt, das dem Wert der Variablen am Beginn der Seite entspricht. In den Fußzeilen wird das Wertelabel angezeigt, das dem Wert der Variablen am Ende der Seite entspricht. Ist kein Wertelabel vorhanden, wird der aktuelle Wert angezeigt.

Sondervariablen. Mit den Sondervariablen *DATE* und *PAGE* können Sie das aktuelle Datum oder die Seitenzahl in eine beliebige Zeile des Kopf- oder Fußzeilenbereichs des Berichts eingeben. Wenn Ihre Datendatei Variablen wie *DATE* oder *PAGE* enthält, können Sie diese in Titeln oder Fußzeilen des Berichts nicht verwenden.

Bericht in Spalten

Mit “Bericht in Spalten” werden Auswertungsberichte erstellt, die in verschiedenen Spalten unterschiedliche Auswertungsstatistiken enthalten.

Beispiel. In einem Einzelhandelsunternehmen mit Filialen werden Informationen über Angestellte, Gehälter, Anstellungszeiten sowie Filiale und Abteilung jedes Beschäftigten in Datensätzen gespeichert. Sie können einen Bericht erstellen, der eine zusammenfassende Gehaltsstatistik (zum Beispiel Mittelwert, Minimum und Maximum) für jede Abteilung liefert.

Datenspalten. Hier werden die Berichtsvariablen aufgelistet, für die Sie eine Auswertungsstatistik anfordern möchten, und das Anzeigeformat sowie die für jede Variable angezeigte Auswertungsstatistik festgelegt.

Break-Spalten. Hiermit werden optionale Break-Variablen, die den Bericht in Gruppen aufteilen, aufgelistet und das Anzeigeformat der Break-Spalten festgelegt. Bei mehreren Break-Variablen wird für jede Kategorie einer Break-Variablen eine getrennte Gruppe innerhalb der Kategorien der vorhergehenden Break-Variablen in der Liste erzeugt. Die Break-Variablen müssen diskrete kategoriale Variablen sein, welche die Fälle in eine begrenzte Anzahl von sinnvollen Kategorien aufteilen.

Bericht. Hiermit legen Sie alle Merkmale des Berichts fest, beispielsweise die Anzeige der fehlenden Werte, Seitennumerierung und Titel.

Vorschau. Es wird nur die erste Seite des Berichtes angezeigt. Mit dieser Option erhalten Sie eine Vorschau auf das Format Ihres Berichts, ohne diesen komplett bearbeiten zu müssen.

Daten sind schon sortiert. Bei Berichten mit Break-Variablen muß die Datendatei vor dem Erstellen des Berichts nach den Werten der Break-Variablen sortiert werden. Wenn Ihre Datendatei bereits nach den Werten der Break-Variablen sortiert ist, können Sie durch Auswahl dieser Option Bearbeitungszeit einsparen. Diese Option ist besonders hilfreich, wenn Sie bereits einen Bericht für die Vorschau erstellt haben.

Beispielausgabe

Abbildung 37-9
Auswertungsbericht mit Auswertungsstatistik in Spalten

Division	Mean Annual Mean Age	Minimum Annual Salary	Maximum Annual Salary	Salary
Carpeting	30.75	$11,754	$9,200	$19,500
Appliances	31.11	$12,508	$7,500	$28,300
Furniture	36.87	$13,255	$8,975	$17,050
Hardware	36.20	$17,580	$7,450	$22,500

So erstellen Sie eine Zusammenfassung: Bericht in Spalten

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Berichte
  Bericht in Spalten...

- Wählen Sie mindestens eine Variable für die Datenspalten aus. Für jede ausgewählte Variable wird eine Spalte im Bericht erzeugt.

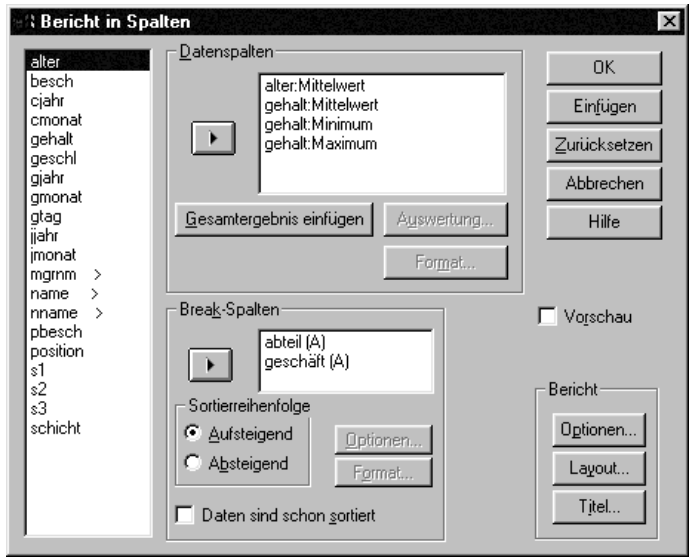
- Um das Auswertungsmaß für eine Variable zu ändern, wählen Sie die Variable in der Liste “Datenspalten” aus, und klicken Sie auf Auswertung.

- Um mehr als ein Auswertungsmaß für eine Variable berechnen zu lassen, wählen Sie die Variable in der Quelliste aus und übernehmen diese für jedes gewünschte Auswertungsmaß in die Liste “Datenspalten”.

- Um eine Spalte mit Summe, Mittelwert, Verhältnis oder einer anderen Funktion einer vorhandenen Spalte anzuzeigen, klicken Sie auf Gesamtergebnis einfügen. Dadurch wird die Variable *Gesamt* in die Liste “Datenspalten” aufgenommen.

- Wählen Sie bei sortierten und nach Untergruppen angezeigten Berichten mindestens eine Variable für die Break-Spalten aus.

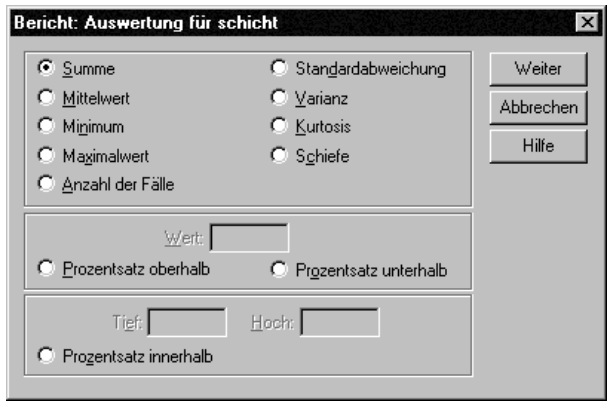
Abbildung 37-10
Dialogfeld "Bericht in Spalten"



Datenspalten: Auswertungsfunktion

Im Dialogfeld "Auswertung" wird die angezeigte Auswertungsstatistik der ausgewählten Datenspalten-Variablen festgelegt.

Abbildung 37-11
Dialogfeld "Bericht: Auswertung"



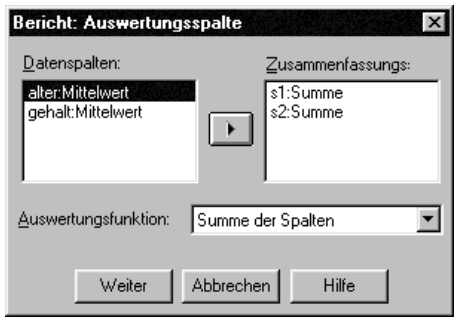
Die verfügbaren Auswertungsstatistiken sind Summe, Mittelwert, Minimum, Maximum, Anzahl der Fälle, Prozent der Fälle über oder unter einem festgelegten Wert, Prozent der Fälle innerhalb eines festgelegten Wertebereichs, Standardabweichung, Varianz, Kurtosis und Schiefe.

Auswertungsspalte für Gesamtergebnis

Im Dialogfeld "Bericht: Auswertungsspalte" werden Einstellungen für die Gesamt-Auswertungsstatistik festgelegt, die zwei oder mehr Datenspalten zusammenfaßt.

Die folgenden Gesamt-Auswertungsstatistiken sind verfügbar: Summe der Spalten, Mittelwert der Spalten, Minimum, Maximum, Differenz zwischen den Werten zweier Spalten, Quotient der Werte in einer Spalte dividiert durch die Werte einer anderen Spalte und das Produkt der miteinander multiplizierten Spaltenwerte.

Abbildung 37-12
Dialogfeld "Bericht: Auswertungsspalte"



Summe der Spalten. Die Spalte *Gesamt* enthält die Summe der Spalten in der Liste "Zusammenfassungsspalte".

Mittelwert der Spalten. Die Spalte *Gesamt* enthält den Durchschnitt der Spalten in der Liste "Zusammenfassungsspalte".

Minimum der Spalten. Die Spalte *Gesamt* enthält den Minimalwert der Spalten in der Liste "Zusammenfassungsspalte".

Maximum der Spalten. Die Spalte *Gesamt* enthält den Maximalwert der Spalten in der Liste "Zusammenfassungsspalte".

1. Spalte - 2. Spalte. Die Spalte *Gesamt* enthält die Differenz zwischen den Spalten in der Liste “Zusammenfassungsspalte”. Die Liste “Zusammenfassungsspalte” muß dabei genau zwei Spalten enthalten.

1. Spalte / 2. Spalte. Die Spalte *Gesamt* enthält den Quotienten der Spalten in der Liste “Zusammenfassungsspalte”. Die Liste “Zusammenfassungsspalte” muß dabei genau zwei Spalten enthalten.

% 1. Spalte / 2. Spalte. Die Spalte *Gesamt* enthält den prozentualen Anteil der ersten Spalte an der zweiten Spalte in der Liste “Zusammenfassungsspalte”. Die Liste “Zusammenfassungsspalte” muß dabei genau zwei Spalten enthalten.

Produkt der Spalten. Die Spalte *Gesamt* enthält das Produkt der Spalten in der Liste “Zusammenfassungsspalte”.

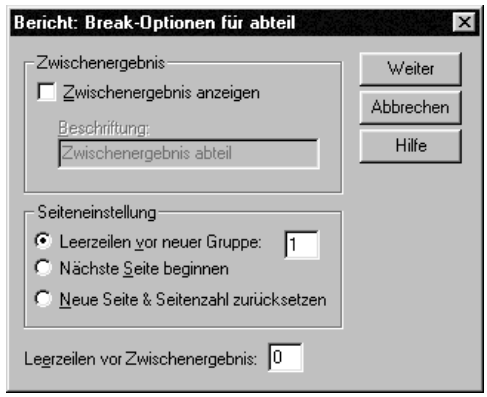
Format der Berichtsspalte

Die Formatoptionen von Daten- und Break-Spalten für “Bericht in Spalten” entsprechen den Optionen für “Bericht in Zeilen”.

Bericht: Break-Optionen für Bericht in Spalten

Mit “Break-Optionen” werden Anzeige der Zwischenergebnisse, Abstand und Seitennumerierung für Break-Kategorien festgelegt.

Abbildung 37-13
Dialogfeld “Bericht: Break-Optionen”



Zwischenergebnis. Hiermit wird die Anzeige der Zwischenergebnisse für Break-Kategorien festgelegt.

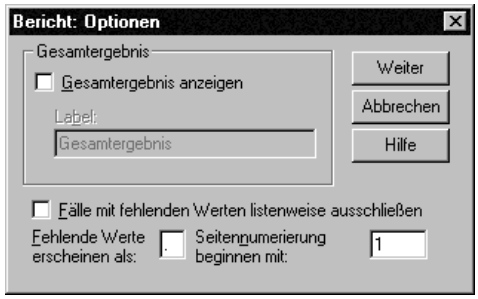
Seiteneinstellung. Hiermit werden Abstand und Seitenaufteilung für Kategorien der ausgewählten Break-Variablen festgelegt. Sie können eine Anzahl von Leerzeilen zwischen den Break-Kategorien festlegen oder eine Break-Kategorie an einen neuen Seitenanfang legen.

Leerzeilen vor Zwischenergebnis. Hiermit legen Sie die Anzahl leerer Zeilen zwischen den Daten der Break-Kategorien und den Zwischenergebnissen fest.

Bericht: Optionen für Bericht in Spalten

Mit "Optionen" werden Anzeige der Gesamtergebnisse, Anzeige der fehlenden Werte und Seitennumerierung in Auswertungsberichten in Spalten festgelegt.

Abbildung 37-14
Dialogfeld "Bericht: Optionen"



Gesamtergebnis. In jeder Spalte wird am unteren Rand ein Gesamtergebnis angezeigt und beschriftet.

Fehlende Werte. Sie können fehlende Werte vom Bericht ausschließen oder fehlende Werte mit einem ausgewählten Zeichen im Bericht kennzeichnen.

Bericht: Layout für Bericht in Spalten

Die Layout-Optionen für "Bericht in Spalten" entsprechen den Optionen für "Bericht in Zeilen".

Zusätzliche Funktionen beim Befehl REPORT

Mit der SPSS-Befehlssprache verfügen Sie über die folgenden zusätzlichen Möglichkeiten:

- In den Spalten einer einzelnen Auswertungszeile lassen sich unterschiedliche Auswertungsfunktionen anzeigen.
- In Datenspalten können Auswertungszeilen für Variablen eingefügt werden, die nicht den Variablen der Datenspalten entsprechen. Außerdem können Zeilen für verschiedene Kombinationen (zusammengesetzte Funktionen) der Auswertungsfunktion eingefügt werden.
- Als Auswertungsfunktionen können Median, Modalwert, Häufigkeit und Prozent verwendet werden.
- Das Anzeigeformat der Auswertungsstatistiken kann genauer festgelegt werden.
- An verschiedenen Stellen des Berichtes können Leerzeilen eingefügt werden.
- In Listenberichten können nach jedem *n*-ten Fall Leerzeilen eingefügt werden.

Wegen der Komplexität der Syntax zum Befehl `REPORT` kann es hilfreich sein, beim Erstellen eines neuen Berichts mit Syntax auf einen vorhandenen Bericht zurückzugreifen. Zum Anpassen eines aus Dialogfeldern erstellten Berichts kopieren Sie die entsprechende Syntax, fügen diese ein und ändern sie so, daß Sie den gewünschten Bericht erstellen können.

Vollständige Informationen zur Syntax finden Sie in der *SPSS Command Syntax Reference*.

Kapitel
38

Reliabilitätsanalyse

Die Reliabilitätsanalyse ermöglicht es Ihnen, die Eigenschaften von Messniveaus und der Items zu untersuchen, aus denen diese sich zusammensetzen. Mit der Prozedur "Reliabilitätsanalyse" können Sie eine Anzahl von allgemein verwendeten Reliabilitäten des Meßniveaus berechnen, und es werden Ihnen Informationen über die Beziehungen zwischen den Items in der Skala zur Verfügung gestellt. Korrelationskoeffizienten in Klassen können verwendet werden, um Reliabilitätsschätzer der Urteiler zu berechnen.

Beispiel. Wird die Kundenzufriedenheit mit Ihrem Fragebogen sinnvoll gemessen? Mit der Reliabilitätsanalyse können Sie das Ausmaß des Zusammenhangs zwischen den Items in Ihrem Fragebogen bestimmen, einen globalen Index der Reproduzierbarkeit bzw. der inneren Konsistenz der vollständigen Skala ermitteln und die kritischen Items herausfinden, welche nicht mehr in der Skala verwendet werden sollten.

Statistiken. Deskriptive Statistiken für jede Variable und für die Skala, Auswertungsstatistik für mehrere Items, Inter-Item-Korrelationen und Inter-Item-Kovarianzen, Reliabilitätsschätzer, ANOVA-Tabelle, Korrelationskoeffizient in Klassen, T^2 nach Hotelling und Tukey-Additivitätstest.

Modelle. Die folgenden Reliabilitätsmodelle sind verfügbar:

- **Alpha (Cronbach).** Dieses Modell ist ein Modell der inneren Konsistenz, welches auf der durchschnittlichen Inter-Item-Korrelation beruht.
- **Split-Half.** Bei diesem Modell wird die Skala in zwei Hälften geteilt und die Korrelation zwischen den Hälften berechnet.
- **Guttman.** Bei diesem Modell werden Guttmans untere Grenzen für die wahre Reliabilität berechnet.

- **Parallel.** Bei diesem Modell wird angenommen, daß alle Items gleiche Varianzen und gleiche Fehlervarianzen für mehrere Wiederholungen aufweisen.
- **Streng parallel.** Bei diesem Modell gelten die Annahmen des parallelen Modells, und es wird zusätzlich die Gleichheit der Mittelwerte der Items angenommen.

Daten. Die Daten können dichotom, ordinal- oder intervallskaliert sein. Sie müssen jedoch numerisch kodiert sein.

Annahmen. Die Beobachtungen sollten unabhängig sein, und Fehler dürfen zwischen den Items nicht korrelieren. Jedes Paar von Items sollte bivariat normalverteilt sein. Die Skalen sollten additiv sein, so daß sich jedes Item linear zum Gesamtwert verhält.

Verwandte Prozeduren. Wenn Sie die Dimensionalität der Skalen-Items untersuchen möchten (um herauszufinden, ob mehr als eine Konstruktion nötig ist, um das Muster der Item-Werte zu erklären), verwenden Sie die Prozedur "Faktorenanalyse" oder "Multidimensionale Skalierung". Wenn Sie homogene Variablengruppen identifizieren möchten, verwenden Sie die Prozedur "Hierarchische Clusteranalyse", um Variablen zu clustern.

So lassen Sie eine Reliabilitätsanalyse berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

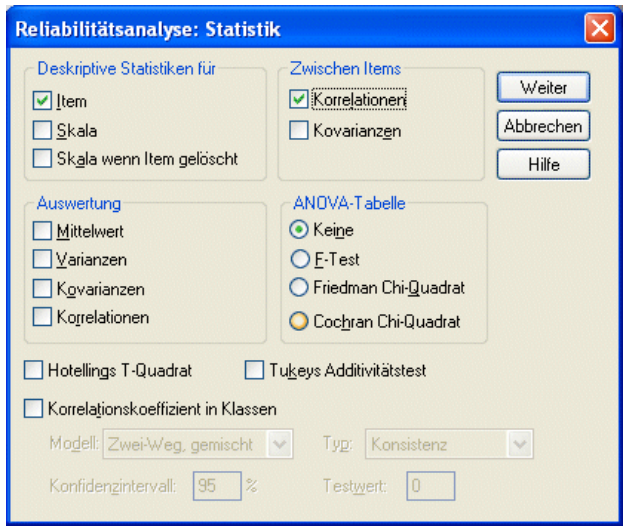
Analysieren
  Skala
    Reliabilitätsanalyse...

Abbildung 38-1
Dialogfeld "Reliabilitätsanalyse"

- Wählen Sie mindestens zwei Variablen als potentielle Komponenten einer additiven Skala aus.
- Wählen Sie aus dem Dropdown-Listenfeld “Modell” ein Modell aus.

Reliabilitätsanalyse: Statistik

Abbildung 38-2
Dialogfeld “Reliabilitätsanalyse: Statistik”



Sie können zahlreiche Statistiken auswählen, die sowohl die Skala als auch die Items beschreiben. Die Statistiken, die in der Standardeinstellung angezeigt werden, umfassen die Anzahl der Fälle, die Anzahl der Items und die folgenden Reliabilitätsschätzer:

- Alpha-Modelle: Alpha-Koeffizient. Bei dichotomen Daten entspricht dieser dem Kuder-Richardson-20-(KR20-)Koeffizienten.
- Split-Half-Modelle: Korrelation zwischen den beiden Hälften, Split-Half-Reliabilität nach Guttman, Spearman-Brown-Reliabilität (gleiche und ungleiche Länge) und Alpha-Koeffizienten für jede Hälfte.

- Guttman-Modelle: Reliabilitätskoeffizienten Lambda 1 bis Lambda 6.
- Parallele und streng parallele Modelle: Anpassungstest für das Modell, Schätzer der Fehlervarianz, der Gesamtvarianz und der wahren Varianz, geschätzte gemeinsame Inter-Item-Korrelation, geschätzte Reliabilität und unverzerrter Schätzer der Reliabilität.

Deskriptive Statistiken für. Erzeugt deskriptive Statistiken für Skalen oder Items über Fälle.

- **Item.** Erzeugt deskriptive Statistiken für Items über Fälle.
- **Skala.** Erzeugt deskriptive Statistiken für Skalen.
- **Skala, wenn Item gelöscht.** Zeigt die Auswertungsstatistik an, bei der jedes Item mit der Skala verglichen wird, die aus den anderen Items aufgebaut wurde. Zu den statistischen Angaben gehören auch Mittelwert und Varianz der Skala, falls das Item aus der Skala gelöscht würde, die Korrelation zwischen dem Element und der Skala aus den anderen Items sowie Cronbachs Alpha, falls das Element aus der Skala gelöscht würde.

Auswertung. Hiermit werden deskriptive Statistiken der Item-Verteilungen für alle Items in der Skala berechnet.

- **Mittelwerte (Reliability).** Statistik für die Item-Mittelwerte. Angezeigt werden der kleinste, größte und durchschnittliche Item-Mittelwert, der Bereich und die Varianz der Item-Mittelwerte sowie das Verhältnis zwischen dem größten und dem kleinsten Item-Mittelwert.
- **Varianzen.** Auswertungsstatistik für Varianzen der Items. Es werden die kleinsten, größten und mittleren Varianzen der Items, die Spannweite und die Varianz der Item-Varianzen sowie das Verhältnis zwischen der größten und der kleinsten Varianzen angezeigt.
- **Kovarianzen.** Statistik für die Kovarianzen zwischen den Items. Von den Kovarianzen zwischen den Items werden der kleinste und der größte Wert, der Mittelwert, die Spannweite und die Varianz, sowie das Verhältnis vom größten zum kleinsten Wert angezeigt.
- **Korrelationen.** Statistik für die Korrelationen zwischen den Items. Von den Korrelationen zwischen den Items werden der kleinste und der größte Wert, der Mittelwert, die Spannweite und die Varianz, sowie das Verhältnis vom größten zum kleinsten Wert angezeigt.

Inter-Item. Hiermit werden Matrizen der Korrelationen oder Kovarianzen zwischen den Items erstellt.

ANOVA-Tabelle. Hiermit werden Tests auf gleiche Mittelwerte berechnet.

- **F-Test.** Zeigt eine Tabelle zur Varianzanalyse mit Meßwiederholtungen an.
- **Friedman Chi-Quadrat.** Zeigt das Chi-Quadrat nach Friedman und den Konkordanz-Koeffizienten nach Kendall an. Diese Option ist für Daten geeignet, die in Form von Rängen vorliegen. Der Chi-Quadrat-Test ersetzt den üblichen F-Test in der ANOVA-Tabelle.
- **Cochran Chi-Quadrat.** Zeigt Cochrans Q-Test an. Diese Option ist für dichotome Daten geeignet. Die Q-Statistik ersetzt die übliche F-Statistik in der ANOVA-Tabelle.

Hotellings T-Quadrat. Erzeugt einen multivariaten Test der Nullhypothese, daß alle Items auf der Skala den gleichen Mittelwert besitzen.

Tukeys Additivitätstest. Erzeugt einen Test der Annahme, daß zwischen den Items keine multiplikative Wechselwirkung besteht.

Korrelationskoeffizienten in Klassen. Erzeugt ein Maß der Konsistenz oder Werteübereinstimmung innerhalb von Fällen.

- **Modell.** Wählen Sie das Modell für die Berechnung des Korrelationskoeffizienten in Klassen. Verfügbar sind die Modelle "Zwei-Weg, gemischt", "Zwei-Weg, zufällig" und "Ein-Weg, zufällig". Wählen Sie Zwei-Weg, gemischt aus, wenn die Personeneffekte zufällig und die Item-Effekte fest sind. Wählen Sie Zwei-Weg, zufällig aus, wenn die Personeneffekte und die Item-Effekte zufällig sind. Wählen Sie Ein-Weg, zufällig aus, wenn die Personeneffekte zufällig sind.
- **Typ.** Wählen Sie den Indextyp. "Konsistenz" und "Absolute Übereinstimmung" sind verfügbar.
- **Konfidenzintervall.** Legen Sie das Niveau des Konfidenzintervalls fest. Der Standardwert ist 95%.
- **Testwert.** Legen Sie den hypothetischen Wert des Koeffizienten für den Hypothesentest fest. Dies ist der Wert, mit dem der beobachtete Wert verglichen wird. Der Standardwert ist 0.

Zusätzliche Funktionen beim Befehl RELIABILITY

Mit der SPSS-Befehlssprache verfügen Sie über die folgenden zusätzlichen Möglichkeiten:

- Korrelationsmatrizen können gelesen und analysiert werden.
- Korrelationsmatrizen können für spätere Analysen gespeichert werden.
- Für die Split-Half-Methode können Aufteilungen festgelegt werden, die nicht genau Hälften entsprechen.

Vollständige Informationen zur Syntax finden Sie im *SPSS Syntax Reference Guide*.

Kapitel
39

Multidimensionale Skalierung

Bei der multidimensionalen Skalierung wird versucht, die Struktur in einem Set von Distanzmaßen zwischen Objekten oder Fällen zu erkennen. Diese Aufgabe wird durch das Zuweisen von Beobachtungen zu bestimmten Positionen in einem konzeptuellen Raum (gewöhnlich zwei- oder dreidimensional) erzielt, und zwar so, dass die Distanzen zwischen den Punkten des Raums mit den gegebenen Unähnlichkeiten so gut wie möglich übereinstimmen. In vielen Fällen können die Dimensionen dieses konzeptuellen Raums interpretiert und für ein besseres Verständnis Ihrer Daten verwendet werden.

Wenn Sie über objektiv gemessene Variablen verfügen, können Sie die multidimensionale Skalierung als Technik zur Datenreduktion verwenden (erforderlichenfalls berechnet die Prozedur "Multidimensionale Skalierung" die Distanzen aus multivariaten Daten für Sie). Die multidimensionale Skalierung kann auch auf subjektive Einschätzungen von Unähnlichkeiten zwischen Objekten oder Konzepten angewendet werden. Außerdem kann die Prozedur "Multidimensionale Skalierung" Unähnlichkeitsdaten aus mehreren Quellen verarbeiten, beispielsweise von mehreren Befragern oder Befragten einer Umfrage.

Beispiel. Wie nehmen Personen Ähnlichkeiten zwischen unterschiedlichen Autos wahr? Wenn Sie über Daten verfügen, in denen Befragte ihre Einschätzungen der Ähnlichkeiten von verschiedenen Automarken und -modellen abgegeben haben, kann die multidimensionale Skalierung zur Identifizierung der Dimensionen verwendet werden, welche die Wahrnehmungen von Käufern beschreibt. Sie könnten zum Beispiel feststellen, dass Preis und Größe eines Fahrzeuges einen zweidimensionalen Raum definieren, welcher die von den Befragten geäußerten Ähnlichkeiten erklärt.

Statistiken. Für jedes Modell: Datenmatrix, optimal skalierte Datenmatrix, S-Stress (Young), Stress (Kruskal), RSQ, Stimulus-Koordinaten, durchschnittlicher Stress und RSQ für jeden Stimulus (RMDS-Modelle). Für Modelle der individuellen Differenzen (INDSCAL): Subjektgewichtungen und Seltsamkeits-Index ("weirdness index") für jedes Subjekt. Für jede Matrix in replizierten Modellen für die multidimensionale

Skalierung: Stress und RSQ für jeden Stimulus. Diagramme: Stimulus-Koordinaten (zwei- oder dreidimensional), Streudiagramm von Unähnlichkeiten über Distanzen.

Daten. Wenn Sie über Unähnlichkeitsdaten verfügen, sollten alle Unähnlichkeiten quantitativ und mit derselben Maßeinheit gemessen sein. Wenn Sie über multivariate Daten verfügen, können die Variablen quantitativ, binär oder Häufigkeitsdaten sein. Die Skalierung der Variablen ist ein wichtiger Punkt. Unterschiede in der Skalierung können Ihre Lösung beeinflussen. Wenn Ihre Variablen große Differenzen in der Skalierung aufweisen (wenn zum Beispiel eine Variable in Dollar und die andere Variable in Jahren gemessen wird), sollten Sie deren Standardisierung in Betracht ziehen (dies kann mit der Prozedur "Multidimensionale Skalierung" automatisch durchgeführt werden).

Annahmen. Die Prozedur "Multidimensionale Skalierung" ist relativ frei von Annahmen zur Verteilung. Stellen Sie sicher, dass Sie im Dialogfeld "Multidimensionale Skalierung: Optionen" ein geeignetes Meßniveau auswählen (Ordinal-, Intervall- oder Verhältnisdaten), sodass Ihre Ergebnisse richtig berechnet werden können.

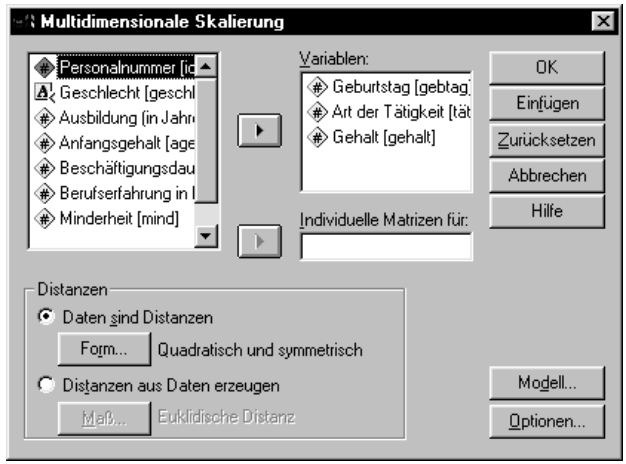
Verwandte Prozeduren. Wenn Sie eine Datenreduktion durchführen möchten, können Sie auch eine Faktoranalyse durchführen, insbesondere bei quantitativen Variablen. Wenn Sie Gruppen von ähnlichen Fällen identifizieren möchten, können Sie die multidimensionale Skalierung durch eine hierarchische Clusteranalyse oder eine Clusterzentrenanalyse ergänzen.

So berechnen Sie eine multidimensionale Skalierung:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Skala
    Multidimensionale Skalierung...

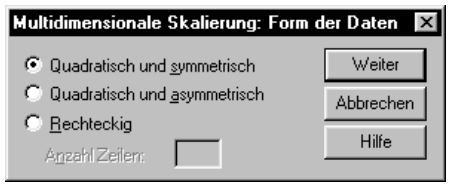
Abbildung 39-1
Dialogfeld "Multidimensionale Skalierung"



- Wählen Sie unter "Distanzen" entweder Daten sind Distanzen oder Distanzen aus Daten erzeugen aus.
- Wenn Ihre Daten Distanzen darstellen, wählen Sie mindestens vier numerische Variablen für die Analyse aus. (Sie können auch auf Form klicken, um die Form der Distanzmatrix anzugeben.)
- Wenn Sie möchten, dass SPSS die Distanzen erstellt, bevor diese analysiert werden, wählen Sie mindestens eine numerische Variable aus. (Sie können auch auf Maß klicken, um den Typ des gewünschten Distanzmaßes anzugeben.) Jede Gruppenvariable kann entweder numerisch oder ein String sein, und Sie können getrennte Matrizen für jede Kategorie einer Gruppenvariablen erstellen, indem Sie diese Variable in die Liste "Individuelle Matrizen für" verschieben.

Multidimensionale Skalierung: Form der Daten

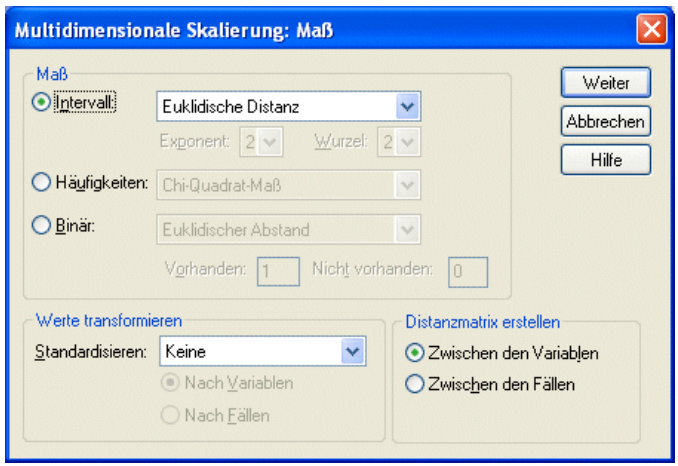
Abbildung 39-2
Dialogfeld "Multidimensionale Skalierung: Form der Daten"



Wenn die Arbeitsdatei Distanzen innerhalb einer Gruppe von Objekten oder zwischen zwei Gruppen von Objekten darstellt, müssen Sie die Form der Datenmatrix angeben, um die richtigen Ergebnisse zu erhalten. *Anmerkung*: Sie können Quadratisch und symmetrisch nicht auswählen, wenn im Dialogfeld "Modell" eine Konditionalität der Zeilen festgelegt ist.

Multidimensionale Skalierung: Distanzen aus Daten erstellen

Abbildung 39-3
Dialogfeld "Multidimensionale Skalierung: Distanzen aus Daten erstellen"



Die multidimensionale Skalierung verwendet Unähnlichkeitsdaten, um eine Skalierungslösung zu erstellen. Wenn Ihre Daten multivariate Daten darstellen (Werte gemessener Variablen), müssen Sie Unähnlichkeitsdaten erstellen, um eine

multidimensionale Skalierungslösung berechnen zu können. Sie können Optionen für das Erstellen von Unähnlichkeitsmaßen aus Ihren Daten festlegen.

Maß. Hier können Sie das Unähnlichkeitsmaß für Ihre Analyse festlegen. Wählen Sie im Gruppenfeld "Maß" die Option aus, die Ihrem Datentyp entspricht. Wählen Sie dann aus dem Dropdown-Listenfeld ein Maß aus, das diesem Messwerttyp entspricht. Die folgenden Optionen sind verfügbar:

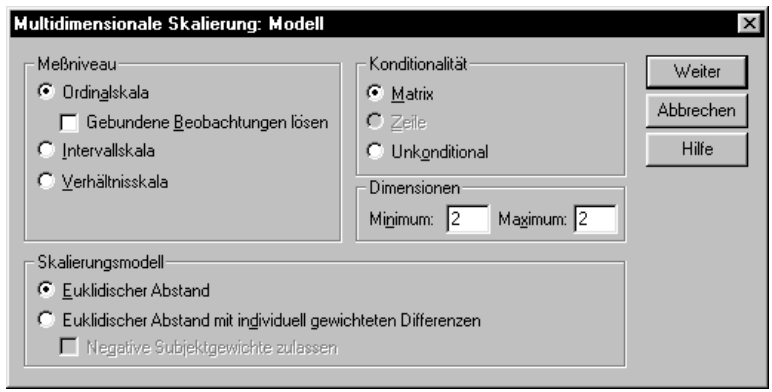
- **Intervall**. Euklidischer Abstand, quadrierter Euklidischer Abstand, Tschebyscheff, Block, Minkowski oder ein benutzerdefiniertes Maß.
- **Anzahl**. Chi-Quadrat-Maß oder Phi-Quadrat-Maß.
- **Binär**. Euklidischer Abstand, quadrierter Euklidischer Abstand, Größendifferenz, Musterdifferenz, Varianz und Distanzmaß nach Lance und Williams.

Distanzmatrix erstellen. Mit dieser Funktion können Sie die Einheit der Analyse wählen. Zur Auswahl stehen "Zwischen den Variablen" oder "Zwischen den Fällen".

Werte transformieren. In bestimmten Fällen, zum Beispiel wenn die Variablen mit sehr unterschiedlichen Skalen gemessen werden, empfiehlt sich das Standardisieren der Werte vor dem Berechnen der Ähnlichkeiten (nicht auf binäre Daten anwendbar). Wählen Sie in der Dropdown-Liste "Standardisieren" eine Standardisierungsmethode aus. Wenn keine Standardisierung erforderlich ist, wählen Sie Keine aus.

Multidimensionale Skalierung: Modell

Abbildung 39-4
Dialogfeld "Multidimensionale Skalierung: Modell"



Die richtige Schätzung eines Modells für die multidimensionale Skalierung hängt von Aspekten der Daten und dem Modell selbst ab.

Meßniveau. Mit dieser Funktion können Sie das Niveau Ihrer Daten festlegen. Die Optionen "Ordinalskala", "Intervallskala" und "Verhältnisskala" sind verfügbar. Wenn die Variablen ordinal sind, können Sie Gebundene Beobachtungen lösen auswählen. Die Variablen werden dann wie stetige Variablen behandelt, sodass die Bindungen (gleiche Werte für unterschiedliche Fälle) optimal gelöst werden können.

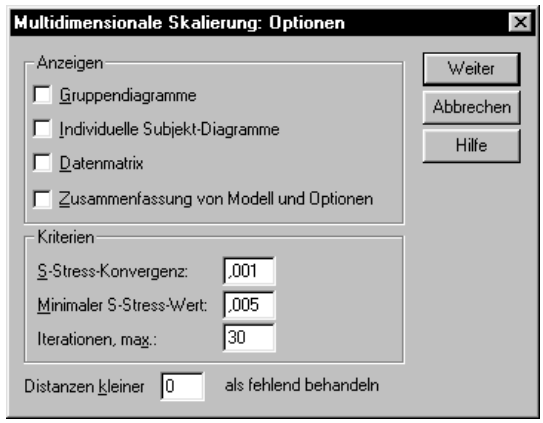
Konditionalität. Hiermit können sie festlegen, welche Vergleiche sinnvoll sind. Als Optionen sind "Matrix", "Zeile" und "Unkonditional" verfügbar.

Dimensionen. Mit dieser Funktion können Sie die Anzahl der Dimensionen für die Skalierungslösung(en) festlegen. Für jede Zahl im Bereich wird eine Lösung berechnet. Legen Sie ganze Zahlen zwischen 1 und 6 fest. Ein Minimum von 1 ist nur möglich, wenn Sie als Skalierungsmodell Euklidischer Abstand auswählen. Legen Sie die gleiche Zahl für das Minimum und das Maximum fest, wenn Sie nur eine Lösung wünschen.

Skalierungsmodell. Hiermit können Sie die Annahmen festlegen, nach denen die Skalierung durchgeführt wird. Als Optionen sind "Euklidischer Abstand" oder "Euklidischer Abstand mit individuell gewichteten Differenzen" (auch als INDSCAL bekannt) verfügbar. Beim Modell "Euklidischer Abstand mit individuell gewichteten Differenzen" können Sie Negative Subjektgewichte zulassen auswählen, wenn dies für Ihre Daten geeignet ist.

Multidimensionale Skalierung: Optionen

Abbildung 39-5
Dialogfeld "Multidimensionale Skalierung: Optionen"



Sie können Optionen für die Analyse der multidimensionalen Skalierung festlegen.

Anzeigen. Mit dieser Funktion können Sie verschiedene Ausgabetypen auswählen. Die Optionen "Gruppendiagramme", "Individuelle Subjekt-Diagramme", "Datenmatrix" und "Zusammenfassung von Modell und Optionen" sind verfügbar.

Kriterien. Hiermit können Sie bestimmen, wann die Iterationen beendet werden sollen. Um die Standardeinstellungen zu ändern, geben Sie Werte für S-Stress-Konvergenz, Minimaler S-Stress-Wert und Iterationen, max. ein.

Distanzen kleiner n als fehlend behandeln. Distanzen, die einen geringeren Wert als diesen Wert aufweisen, werden aus der Analyse ausgeschlossen.

Zusätzliche Funktionen beim Befehl ALSCAL

Mit der SPSS-Befehlssprache verfügen Sie über die folgenden zusätzlichen Möglichkeiten:

- Es können drei weitere Modelltypen verwenden werden. Diese sind in der Literatur über die multidimensionale Skalierung als ASCAL, AINDS und GEMSCAL bekannt.
- Es können polynomiale Transformationen von Intervall- und Verhältnisdaten ausgeführt werden.

- Bei ordinalen Daten können statt Distanzen Ähnlichkeiten analysiert werden.
- Es können nominale Daten analysiert werden.
- Verschiedene Koordinatenmatrizen und Gewichtungsmatrizen können in Dateien gespeichert und für eine Analyse erneut eingelesen werden.
- Die multidimensionale Entfaltung kann eingeschränkt werden.

Vollständige Informationen zur Syntax finden Sie im *SPSS Syntax Reference Guide*.

Kapitel
40

Verhältnisstatistik

Die Prozedur "Verhältnisstatistik" bietet eine umfassende Liste mit Auswertungsstatistiken zur Beschreibung des Verhältnisses zwischen zwei metrischen Variablen.

Sie können die Ausgabe nach Werten einer Gruppenvariablen in auf- oder absteigender Reihenfolge sortieren. Der Bericht für die Verhältnisstatistik kann in der Ausgabe unterdrückt werden, und die Ergebnisse können in einer externen Datei gespeichert werden.

Beispiel. Ist das Verhältnis zwischen dem Schätzwert und dem Verkaufspreis von Häusern in fünf Verwaltungsbezirken in etwa gleich? Im Ergebnis der Analyse könnte sich herausstellen, daß die Verteilung der Verhältnisse je nach Bezirk erheblich variiert.

Statistiken. Median, Mittel, gewichtetes Mittel, Konfidenzintervalle, Streuungskoeffizient (COD), medianzentrierter Variationskoeffizient, mittelzentrierter Variationskoeffizient, preisbezogenes Differential (PRD), Standardabweichung, durchschnittliche absolute Abweichung (AAD), Bereich, Mindest- und Höchstwerte sowie der Konzentrationsindex, der für einen benutzerdefinierten Bereich oder Prozentsatz innerhalb des Medianverhältnisses berechnet wird.

Daten. Verwenden Sie zum Kodieren von Gruppenvariablen numerische Codes oder kurze Strings (nominales oder ordinales Niveau der Werte).

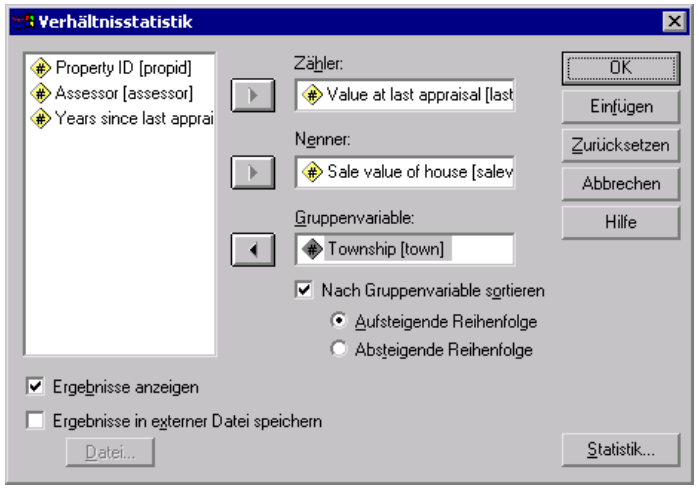
Annahmen. Die Variablen, durch die Zähler und Nenner des Verhältnisses definiert werden, müssen metrische Variablen sein, die positive Werte akzeptieren.

So lassen Sie Verhältnisstatistiken berechnen

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Deskriptive Statistiken
    Verhältnis...

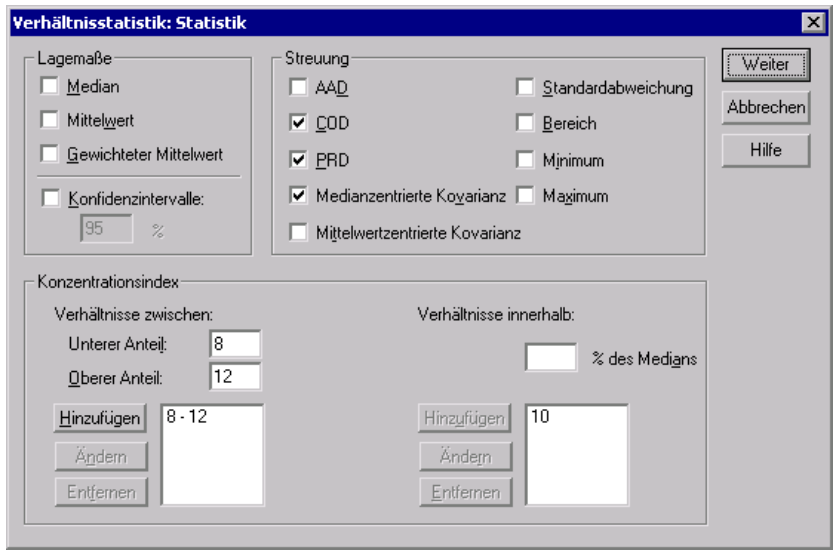
Abbildung 40-1
Dialogfeld "Verhältnisstatistik"



- ▶ Wählen Sie eine Zählervariable.
- ▶ Wählen Sie eine Nennervariable.
- ▶ Die folgenden Optionen sind verfügbar:
  - ■ Wählen Sie eine Gruppenvariable, und legen Sie die Reihenfolge der Gruppen in den Ergebnissen fest.
  - ■ Wählen Sie aus, ob die Ergebnisse im Viewer angezeigt werden sollen.
  - ■ Legen Sie fest, ob die Ergebnisse zur späteren Verwendung in einer externen Datei gespeichert werden sollen, und geben Sie einen Namen für diese Datei an.

Verhältnisstatistik

Abbildung 40-2
Dialogfeld "Verhältnisstatistik"



Lagemaße. Lagemaße sind Statistiken, mit denen die Verteilung von Verhältnissen beschrieben wird.

- **Median.** Der Wert, der sich ergibt, wenn die Anzahl der Verhältnisse unterhalb dieses Werts gleich der Anzahl der Verhältnisse oberhalb dieses Werts ist.
- **Mittelwert.** Das Ergebnis aus der Summierung aller Verhältnisse und der anschließenden Division des Ergebnisses durch die Gesamtanzahl der Verhältnisse.
- **Gewichteter Mittelwert.** Das Ergebnis aus der Division des MittelWerts für den Zähler durch den Mittelwert für den Nenner. Der gewichtete Mittelwert ist außerdem der Mittelwert der durch den Nenner gewichteten Verhältnisse.
- **Konfidenzintervalle.** Mit dieser Option werden Konfidenzintervalle für den Mittelwert, den Median und den gewichteten Mittelwert (falls gewünscht) angezeigt. Geben Sie für das Konfidenzniveau einen Wert größer oder gleich 0 und kleiner als 100 ein.

Streuung. Statistiken, mit denen die Variation oder Streubreite in den beobachteten Werten gemessen wird.

- **AAD.** Die durchschnittliche absolute Abweichung ist die Summe aus den absoluten Abweichungen der Verhältnisse des Medians und der Division des Ergebnisses durch die Gesamtanzahl der Verhältnisse.
- **COD.** Der Streuungskoeffizient entspricht der durchschnittlichen absoluten Abweichung in Prozent des Medians.
- **PRD.** Das preisbezogene Differential, auch Index der Regressivität genannt, ist das Ergebnis der Division des Mittelwerts durch den gewichteten Mittelwert.
- **Medianzentrierte Kovarianz.** Der medianzentrierte Variationskoeffizient entspricht der Wurzel der mittleren quadratischen Abweichung vom Median in Prozent des Medians.
- **Mittelwertzentrierte Kovarianz.** Der mittelwertzentrierte Variationskoeffizient entspricht der Standardabweichung in Prozent des Mittelwerts.
- **Standardabweichung.** Die Standardabweichung ist das Ergebnis der Summierung der quadratischen Abweichungen der Verhältnisse zum Mittelwert, der Division des Ergebnisses durch die Gesamtanzahl der Verhältnisse minus eins und der Berechnung der positiven Quadratwurzel.
- **Bereich.** Der Bereich ist das Ergebnis der Subtraktion des minimalen Verhältnisses vom maximalen Verhältnis.
- **Minimum.** Das Minimum ist das kleinste Verhältnis.
- **Maximum.** Das Maximum ist das größte Verhältnis.

Konzentrationsindex. Der Konzentrationskoeffizient mißt den prozentualen Anteil der Verhältnisse, die in einem bestimmten Intervall liegen. Dieser Koeffizient kann auf zwei verschiedene Arten berechnet werden:

- **Verhältnisse zwischen.** Bei dieser Option wird das Intervall explizit durch Angabe der unteren und oberen Intervallwerte definiert. Geben Sie Werte für den unteren Anteil und den oberen Anteil ein, und klicken Sie auf Hinzufügen, um ein Intervall auszugeben.
- **Verhältnisse innerhalb.** Bei dieser Option wird das Intervall implizit durch Angabe des prozentualen Medians definiert. Geben Sie einen Wert zwischen 0 und 100 ein, und klicken Sie auf Hinzufügen. Die untere Grenze des Intervalls ist gleich (1 – 0,01 × Wert) × Median. Die obere Grenze ist gleich (1 + 0,01 × Wert) × Median.

Kapitel
41

Übersicht über die Diagrammfunktion

Diagramme mit hoher Auflösung können mit den Verfahren im Menü “Grafiken” und mit etlichen der Verfahren im Menü “Analysieren” erstellt werden. In diesem Kapitel finden Sie eine Übersicht über die Diagrammfunktion.

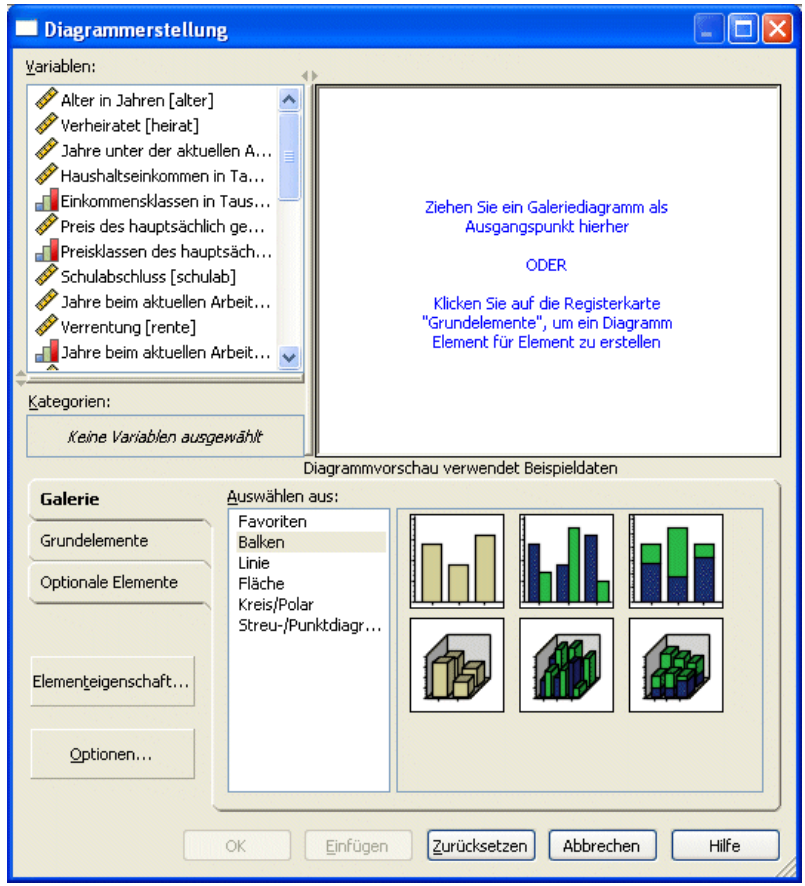
Erstellen und Ändern von Diagrammen

Bevor Sie ein Diagramm erstellen können, müssen Sie über Daten in Ihrem Daten-Editor verfügen. Sie können die Daten direkt in den Daten-Editor eingeben, eine bereits gespeicherte Datendatei öffnen oder ein Arbeitsblatt, eine durch Tabulatoren getrennte Datendatei oder eine Datenbankdatei einlesen. In der Menüauswahl “Lernprogramm” im Menü “Hilfe” finden Sie Online-Beispiele zum Erstellen und Ändern von Diagrammen, und das Online-Hilfesystem bietet Informationen darüber, wie alle Diagrammtypen erstellt und geändert werden können.

Erstellen des Diagramms

- Nachdem Sie die Daten im Daten-Editor erfasst haben, wählen Sie aus dem Menü “Grafiken” die Option Diagrammerstellung aus. Damit wird das Dialogfeld “Diagrammerstellung” geöffnet.

Abbildung 41-1
Dialogfeld "Diagrammerstellung"



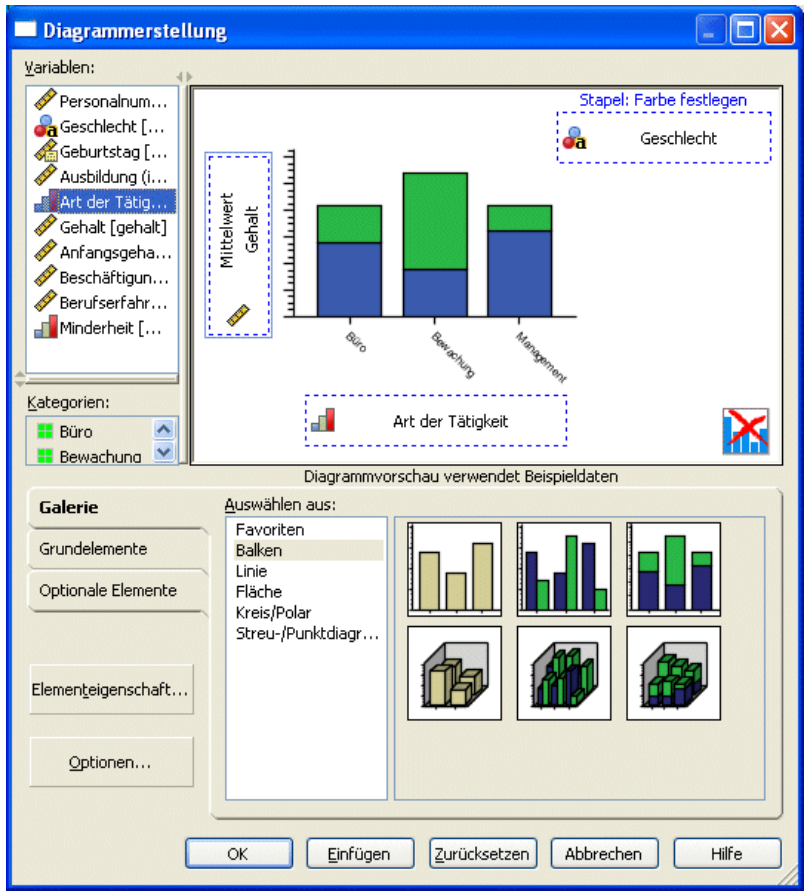
Das Dialogfeld "Diagrammerstellung" wird für häufig eingesetzte Diagrammtypen verwendet, die auf der Registerkarte "Galerie" aufgeführt werden. Wenn Sie ein Diagramm erstellen möchten, das nicht in der Diagrammerstellung verfügbar ist, können Sie auch einen bestimmten Diagrammtyp aus dem Menü "Grafiken" auswählen.

- Ziehen Sie im Dialogfeld "Diagrammerstellung" das Symbol, das dem Diagramm entspricht, auf die Zeichenfläche. Hierbei handelt es sich um die große Fläche oberhalb der Galerie.

- Ziehen Sie die Variablen aus der Liste “Variablen” in die Achsenablagezonen. (Wenn Sie weitere Informationen zum Dialogfeld “Diagrammerstellung” benötigen, klicken Sie auf Hilfe.)

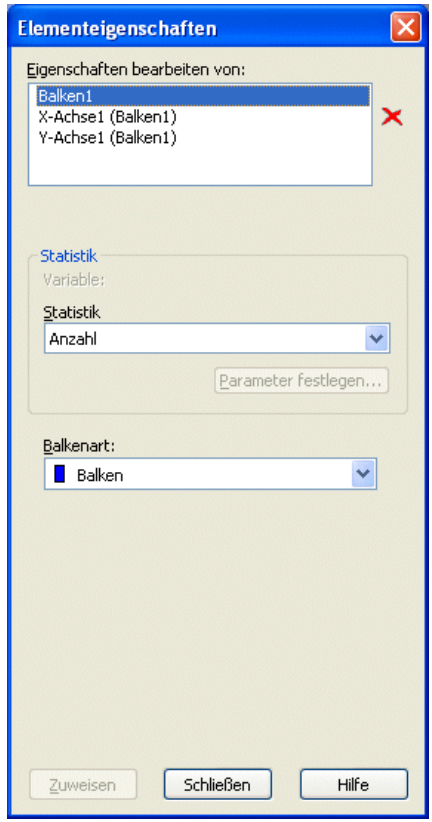
  Wenn das Diagramm vollständig definiert ist, sieht es ähnlich wie das folgende Diagramm aus.

  Abbildung 41-2
  Dialogfeld “Diagrammerstellung” mit belegten Ablagezonen



- Wenn Sie die Statistiken oder die Attribute für die Achsen oder Legenden ändern möchten, klicken Sie auf Elementeigenschaften.

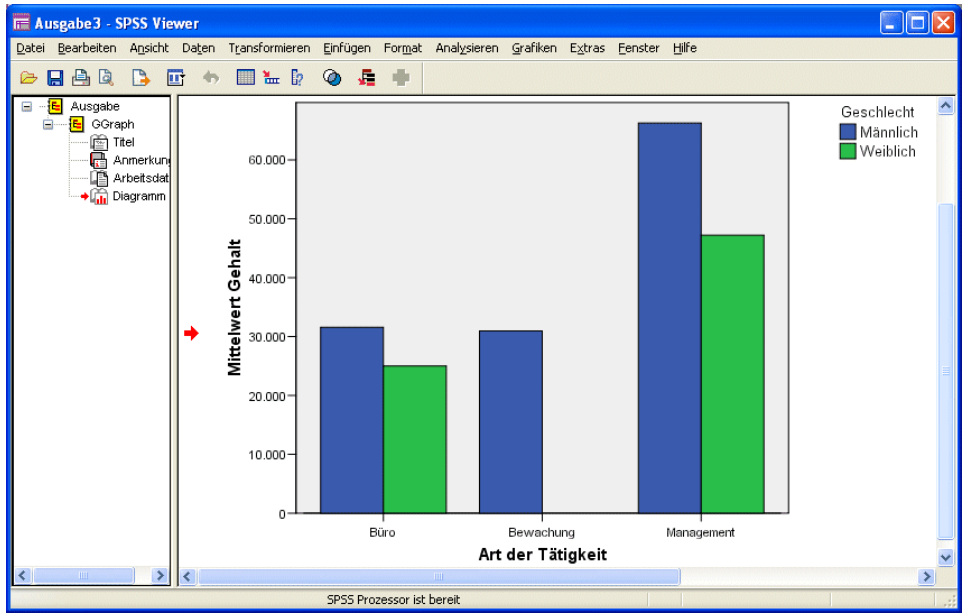
Abbildung 41-3
Fenster "Elementeigenschaften"



- Wählen Sie in der Liste "Eigenschaften bearbeiten von" das Element aus, das Sie bearbeiten möchten. (Wenn Sie Informationen zu den einzelnen Eigenschaften benötigen, klicken Sie auf Hilfe.)

- Klicken Sie im Dialogfeld "Diagrammerstellung" auf OK, um das Diagramm zu erstellen. Das Diagramm wird im Viewer angezeigt.

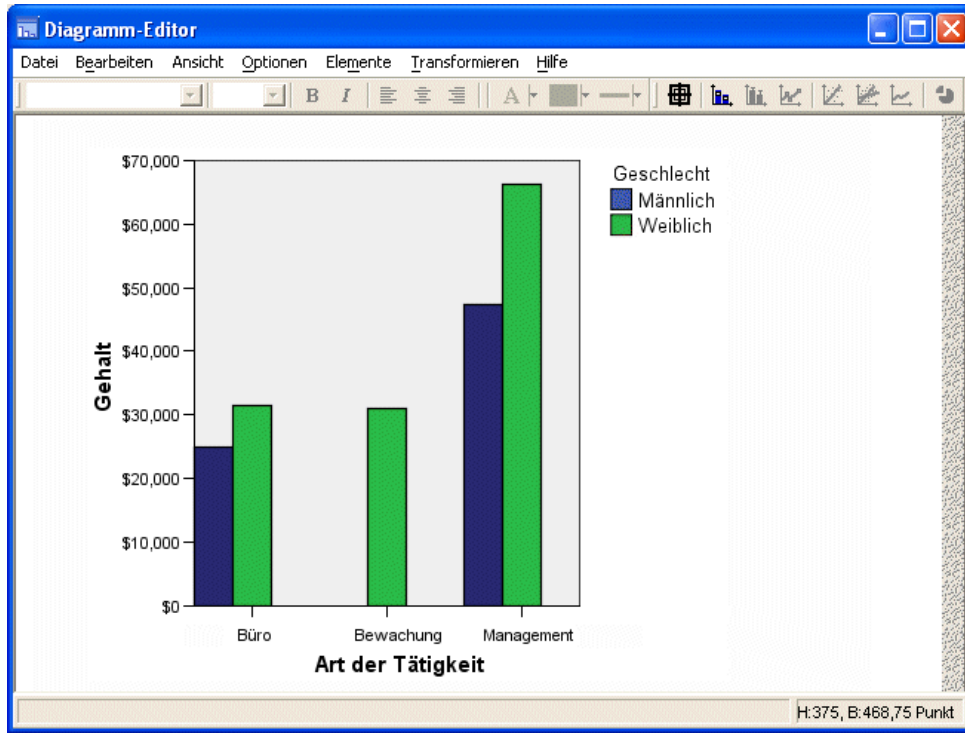
Abbildung 41-4
Balkendiagramm im Viewer-Fenster



Ändern des Diagramms

Doppelklicken Sie zum Ändern eines Diagramms auf eine beliebige Stelle des im Viewer angezeigten Diagramms. Dadurch wird das Diagramm im Diagramm-Editor angezeigt.

Abbildung 41-5
Ursprüngliches Diagramm im Diagrammeditor



Sie können jeden beliebigen Teil des Diagramms ändern oder es in einen anderen Diagrammtyp umwandeln, mit dem dieselben Daten angezeigt werden. Mit den Menüs im Diagramm-Editor können Sie Objekte hinzufügen bzw. sie ein- und ausblenden.

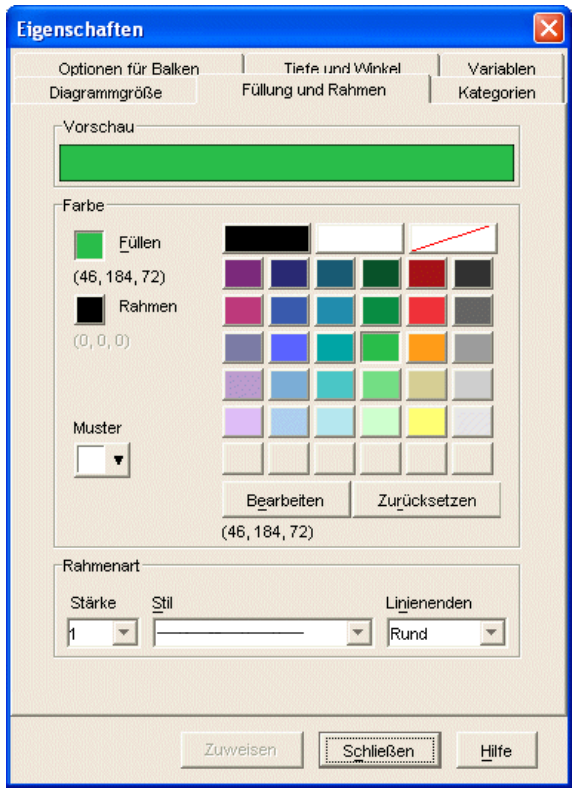
So ändern Sie ein Diagrammobjekt

- Wählen Sie das zu ändernde Objekt aus.
- Wählen Sie die folgenden Befehle aus den Menüs aus:

Bearbeiten
  Eigenschaften...

Dadurch wird das Fenster “Eigenschaften” geöffnet. Die im Fenster “Eigenschaften” angezeigten Registerkarten hängen von Ihrer Auswahl ab. In der Online-Hilfe wird beschrieben, wie Sie die Registerkarten anzeigen können, die Sie benötigen.

Abbildung 41-6
Fenster “Eigenschaften”

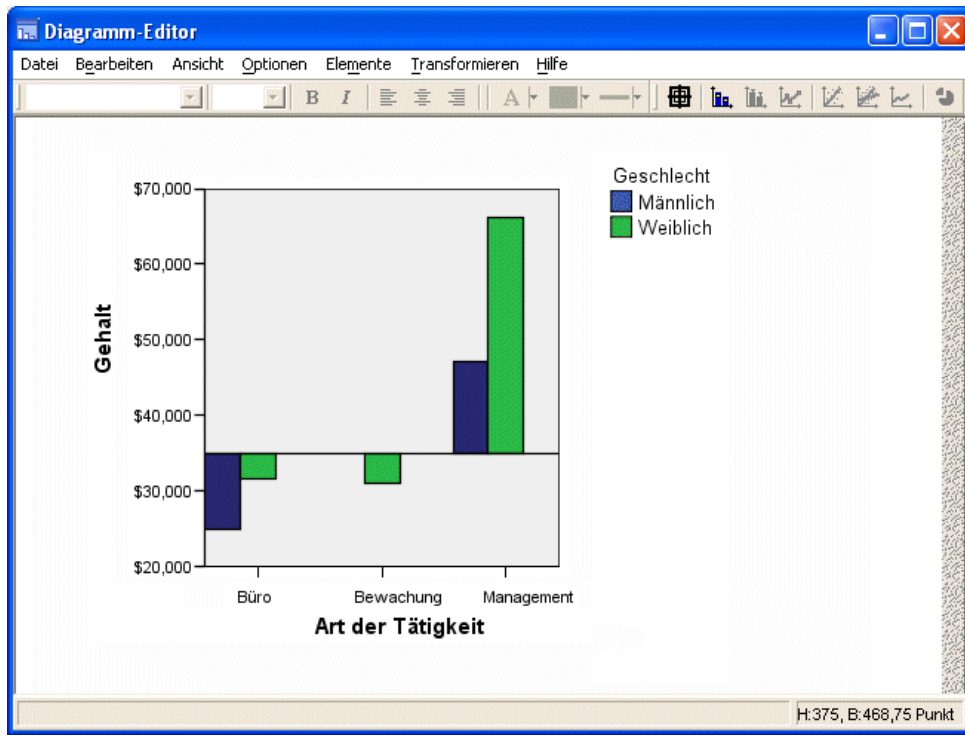


Zu den typischen Änderungen zählen:

- Bearbeiten des Texts im Diagramm
- Ändern der Farbe und des Füllmusters der Balken
- Hinzufügen von Text (z. B. Titel oder Anmerkungen) zum Diagramm
- Ändern der Position der Basislinie für Balken

Im folgenden sehen Sie ein geändertes Diagramm.

Abbildung 41-7
Geändertes Diagramm



Die Änderungen am Diagramm werden beim Schließen des Diagrammfensters gespeichert, und das geänderte Diagramm wird im Viewer angezeigt.

Optionen für die Diagrammdefinition

Wenn Sie ein Diagramm definieren, können Sie Titel hinzufügen und die Optionen für die Diagrammerstellung ändern. Klicken Sie im Dialogfeld “Diagrammerstellung” auf die Registerkarte Optionale Elemente, um Titel, Untertitel und Fußnoten anzugeben. Mit Optionen gelangen Sie zu verschiedenen Diagrammoptionen, z. B. für die Behandlung fehlender Werte oder für die Verwendung von Vorlagen. In den nächsten Abschnitten wird behandelt, wie diese Eigenschaften während des Definierens des Diagramms festgelegt werden.

Titel, Untertitel und Fußnoten

Bei allen Diagrammen können Sie zwei Titelzeilen, eine Untertitelzeile und zwei Fußnotenzeilen als Teil der ursprünglichen Diagrammdefinition festlegen. Um während der Diagrammdefinition Titel oder Fußnoten anzugeben, klicken Sie im Dialogfeld "Diagrammerstellung" auf die Registerkarte Optionale Elemente. Ziehen Sie dann eines der Objekte auf die Zeichenfläche. Sie können den Text ändern, der mit dem Objekt verknüpft ist, indem Sie auf Elementeigenschaften klicken und das Objekt aus der Liste "Eigenschaften bearbeiten von" auswählen.

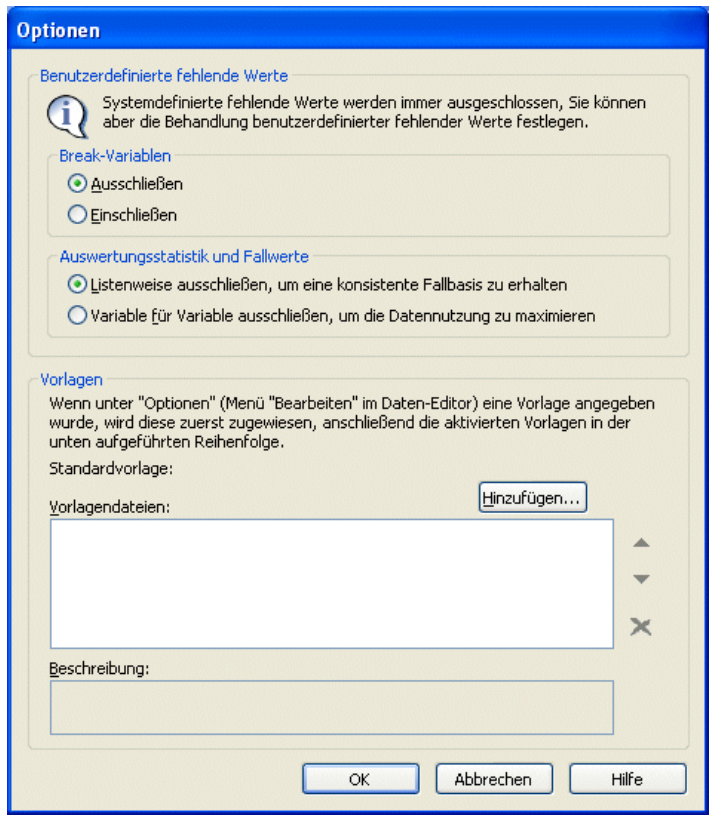
Jede Zeile kann bis zu 72 Zeichen umfassen. Wie viele Zeichen tatsächlich im Diagramm Platz haben, hängt von der Schriftart und der Schriftgröße ab. Die meisten Titel sind standardmäßig linksbündig ausgerichtet und werden, wenn sie zu lang sind, auf der rechten Seite abgeschnitten. Titel von Kreisdiagrammen sind zentriert und werden, wenn sie zu lang sind, an beiden Seiten abgeschnitten.

Titel, Untertitel und Fußnoten werden im Diagramm-Editor als Textfelder wiedergegeben. Sie können Textfelder im Diagramm-Editor hinzufügen, löschen und überarbeiten sowie ihre Schriftart, Schriftgröße und Ausrichtung ändern.

Optionen

Im Dialogfeld "Optionen" finden Sie Optionen für das zu erstellende Diagramm. Sie öffnen das Dialogfeld, indem Sie auf Optionen klicken.

Abbildung 41-8
Dialogfeld "Optionen"


Optionen
Benutzerdefinierte fehlende Werte
Systemdefinierte fehlende Werte werden immer ausgeschlossen, Sie können aber die Behandlung benutzerdefinierter fehlender Werte festlegen.
Break-Variablen
Ausschließen
Einschließen
Auswertungsstatistik und Fallwerte
Listenweise ausschließen, um eine konsistente Fallbasis zu erhalten
Variable für Variable ausschließen, um die Datennutzung zu maximieren
Vorlagen
Wenn unter "Optionen" (Menü "Bearbeiten" im Daten-Editor) eine Vorlage angegeben wurde, wird diese zuerst zugewiesen, anschließend die aktivierten Vorlagen in der unten aufgeführten Reihenfolge.
Standardvorlage:
Vorlagendateien:
Hinzufügen...
Beschreibung:
OK
Abbrechen
Hilfe

Fehlende Werte

Break-Variablen

Wenn in den Daten für die Variablen, die zum Definieren von Kategorien und Untergruppen verwendet werden, fehlende Werte auftreten, wählen Sie Einschließen, um die Kategorie oder Kategorien der benutzerdefinierten fehlenden Werte (Werte, die vom Benutzer als fehlend identifiziert wurden) in das Diagramm einzuschließen. Diese Kategorien verhalten sich bei der Berechnung der Statistik auch als Break-Variablen. Die Kategorien für "Fehlend" werden in der Kategorienachse

oder in der Legende angezeigt. Für diese Kategorien werden einem Diagramm beispielsweise zusätzliche Balken oder Kreissegmente hinzugefügt. Wenn keine fehlenden Werte vorhanden sind, werden die Kategorien für "Fehlend" nicht angezeigt.

Wenn Sie diese Option auswählen und die Anzeige nach dem Erstellen des Diagramms unterdrücken möchten, wählen Sie das Diagramm aus, und wählen Sie anschließend im Menü "Bearbeiten" die Option Eigenschaften aus. Klicken Sie auf die Registerkarte Kategorien, und verschieben Sie die zu unterdrückenden Kategorien in die Liste "Ausgeschlossen".

Auswertungsstatistik und Fallwerte

Sie können eine der folgenden Optionen auswählen, um Fälle mit fehlenden Werten auszuschließen:

- **Listenweise ausschließen, um eine konsistente Fallbasis zu erhalten.** Wenn eine der Variablen im Diagramm einen fehlenden Wert für einen bestimmten Fall aufweist, wird der gesamte Fall aus dem Diagramm ausgenommen.
- **Variable für Variable ausschließen, um die Datennutzung zu maximieren.** Wenn eine ausgewählte Variable fehlende Werte aufweist, werden die Fälle mit diesen fehlenden Werten aus der Analyse der Variablen ausgeschlossen.

Der Unterschied zwischen listenweisem und variablenweisem Ausschluß fehlender Werte wird aus den folgenden Abbildungen ersichtlich, die ein Balkendiagramm für jede der beiden Optionen zeigen.

Abbildung 41-9
Listenweiser Ausschluß fehlender Werte

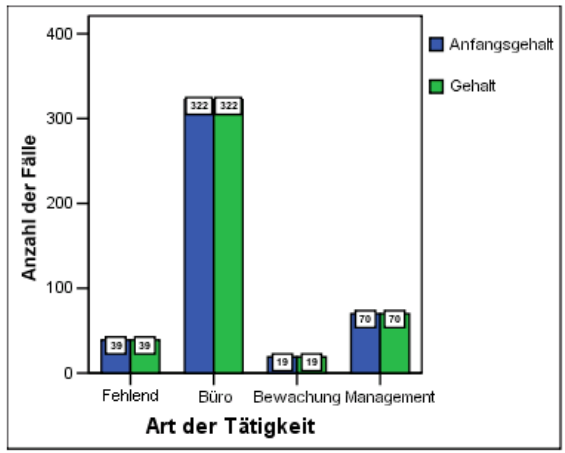
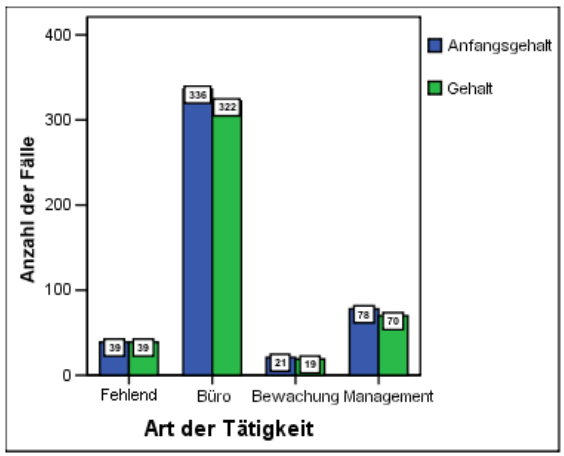


Abbildung 41-10
Ausschluß fehlender Werte Variable für Variable



Die Diagramme wurden aus einer Version der Datei *Employee data.sav* erstellt, die so bearbeitet wurde, daß sie einige systemdefinierte fehlende (leere) Werte in den Variablen für das aktuelle Gehalt und die Art der Tätigkeit aufweist. In anderen Fällen wurde der Wert 0 eingegeben und als fehlend definiert. Bei beiden Diagrammen wurde die Option Fehlende Werte als Kategorie anzeigen ausgewählt, wodurch die Kategorie *Fehlend* zu den anderen angezeigten Kategorien für die Tätigkeitsart

hinzugefügt wird. In jedem Diagramm werden die Werte der Auswertungsfunktion, *Anzahl der Fälle*, in den Balkenbeschriftungen angezeigt.

In beiden Diagrammen weisen 26 Fälle einen systemdefinierten fehlenden Wert für die Art der Tätigkeit auf, und 13 Fälle weisen den benutzerdefinierten fehlenden Wert (0) auf. Im Diagramm für den listenweisen Ausschluß ist die Zahl der Fälle für beide Variablen in jeder Balkengruppe gleich, da bei jedem fehlenden Wert der entsprechende Fall für alle Variablen ausgeschlossen wurde. Im Diagramm für den variablenweisen Ausschluß wird die Anzahl der Fälle ohne fehlende Werte für jede Variable in einer Kategorie dargestellt, ohne die fehlenden Werte in anderen Variablen zu berücksichtigen.

Diagrammvorlagen

Sie können viele Attribute und Textelemente aus einem Diagramm auf ein anderes Diagramm übertragen. Dadurch verfügen Sie über die Möglichkeit, ein Diagramm zu bearbeiten, das Diagramm als Vorlage zu speichern und diese Vorlage anschließend zu verwenden, um eine Reihe ähnlicher Diagramme zu erstellen.

Um bei der Erstellung eines Diagramm eine Vorlage zu verwenden, klicken Sie im Gruppenfeld “Vorlagen” auf Hinzufügen. Dadurch wird ein Dialogfeld zur Auswahl einer Standarddatei geöffnet. Wenn Sie mehrere Vorlagen hinzufügen, werden die Vorlagen in der Reihenfolge angewendet, in der Sie in der Liste “Vorlagendateien” aufgeführt werden. Sie können die Reihenfolge bei Bedarf ändern.

Um eine Vorlage auf ein Diagramm anzuwenden, das sich bereits im Diagramm-Editor befindet, wählen Sie folgende Optionen aus den Menüs aus:

Datei
  Diagrammvorlage zuweisen...

Dadurch wird ein Dialogfeld zur Auswahl einer Standarddatei geöffnet. Wählen Sie eine Datei aus, die als Vorlage verwendet werden soll. Beim Erstellen eines neuen Diagramms wird der ausgewählte Dateiname im Gruppenfeld “Vorlagen” angezeigt, wenn Sie das Dialogfeld für die Diagrammdefinition erneut aufrufen.

Eine Vorlage dient dazu, ein Format aus einem bestehenden Diagramm zu übernehmen und es dem neuen Diagramm, das gerade erstellt wird, zuzuweisen. Im allgemeinen werden alle Formatierungsinformationen aus dem alten Diagramm, die auf das neue Diagramm anwendbar sind, automatisch angewendet. Wenn es sich bei dem alten Diagramm beispielsweise um ein gruppiertes Balkendiagramm handelt, bei dem die Balkenfarben in Gelb und Grün geändert wurden, und das neue Diagramm

ein Mehrfachliniendiagramm ist, werden die Linien gelb und grün. Wenn es sich bei dem alten Diagramm um ein einfaches Balkendiagramm mit Schlagschatten handelt und das neue Diagramm ein einfaches Liniendiagramm ist, erhalten die Linien keine Schlagschatten, da Schlagschatten nicht für Liniendiagramme verwendet werden können. Wenn das Vorlagendiagramm Titel enthält, das neue Diagramm jedoch nicht, werden die Titel aus dem Vorlagendiagramm verwendet. Wenn im neuen Diagramm Titel definiert sind, werden diese verwendet und nicht die Titel aus dem Vorlagendiagramm.

So erstellen Sie eine Diagrammvorlage:

- Erstellen Sie ein Diagramm.
- Bearbeiten Sie das Diagramm so, daß es die für die Vorlage gewünschten Attribute enthält.
- Wählen Sie die folgenden Menübefehle des Diagramm-Editors aus:
  Datei
  Diagrammvorlage speichern...
- Geben Sie im Dialogfeld “Diagrammvorlage speichern” an, welche Eigenschaften des Diagramms in der Vorlage gespeichert werden sollen. In der Online-Hilfe werden die Einstellungen detailliert behandelt.
- Klicken Sie auf Weiter.
- Geben Sie einen Dateinamen und einen Speicherort für die neue Vorlage ein. Die Dateierweiterung der Vorlage lautet *.sgt*.

Kapitel
42

ROC-Kurven

Diese Prozedur stellt einen sinnvollen Weg zur Beurteilung von Klassifikationsschemata dar, bei denen eine Variable mit zwei Kategorien verwendet wird, um Subjekte zu klassifizieren.

Beispiel. Es liegt im Interesse von Banken, Kunden ordnungsgemäß danach zu klassifizieren, ob diese Kunden mit ihren Darlehen in Verzug geraten werden oder nicht. Daher werden spezielle Verfahren für diese Entscheidungen entwickelt. Mit Hilfe von ROC-Kurven kann beurteilt werden, wie gut diese Verfahren funktionieren.

Statistiken. Fläche unter der ROC-Kurve mit Konfidenzintervall und Koordinaten-Punkten der ROC-Kurve. Diagramme: ROC-Kurve.

Methoden. Die Schätzung der Fläche unter der ROC-Kurve kann parameterunabhängig oder parameterabhängig unter Verwendung eines binegativ exponentiellen Modells erfolgen.

Daten. Die Testvariablen sind quantitativ. Die Testvariablen setzen sich oft aus Wahrscheinlichkeiten aus der Diskriminanzanalyse bzw. logistischen Regression zusammen, oder sie werden aus Werten auf einer willkürlichen Skala zusammengesetzt, die anzeigen, wie sehr ein Beurteiler davon "überzeugt" ist, dass ein Subjekt in die eine oder die andere Kategorie fällt. Der Typ der Zustandsvariablen ist nicht vorgegeben. Diese Variable zeigt die tatsächliche Kategorie an, zu der ein Subjekt gehört. Der Wert der Zustandsvariablen zeigt an, welche Kategorie als *positiv* zu betrachten ist.

Annahmen. Es wird angenommen, daß ansteigende Werte auf der Skala des Beurteilers ein Ansteigen der Überzeugung darstellen, daß das Subjekt in die eine Kategorie fällt. Abfallende Werte auf der Skala stellen hingegen eine ansteigende Überzeugung dar, daß das Subjekts der anderen Kategorie angehört. Der Anwender wählt aus, welche Richtung als *positiv* anzusehen ist. Es wird außerdem angenommen, daß die *tatsächliche* Kategorie bekannt ist, zu der jedes Subjekt gehört.

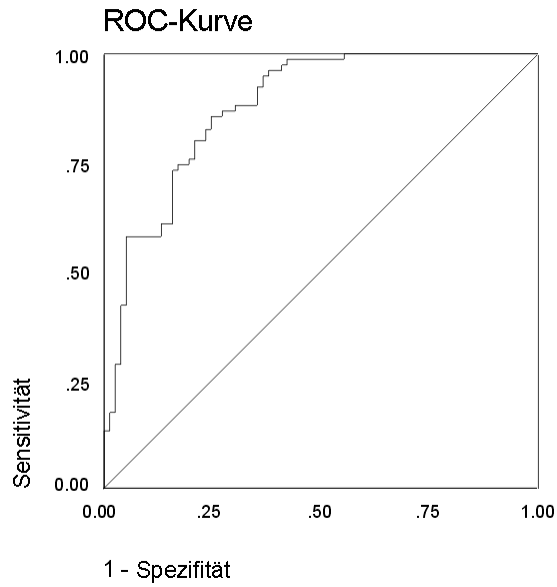
Abbildung 42-1
Ausgabe der ROC-Kurve

Verarbeitete Fälle

Tatsächlich	Gültige N (listenweise)
Positiv [1]	74
Negativ	76

Größere Werte der Variable(n) für das Testergebnis deuten stärker auf einen positiven Ist-Zustand hin.

1. Der positive Ist-Zustand ist 1.00



Fläche unter der Kurve

Variable(n) für Testergebnis: PROBS

Fläche	Standard-fehler[1]	Asymptotische Signifikanz[2]	Asymptotisches 95% Konfidenzintervall	
			Untergrenze	Obergrenze
.877	.028	.000	.823	.931

1. Unter der nichtparametrischen Annahme

2. Nullhypothese: Wahrheitsfläche

So Erstellen Sie eine ROC-Kurve:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

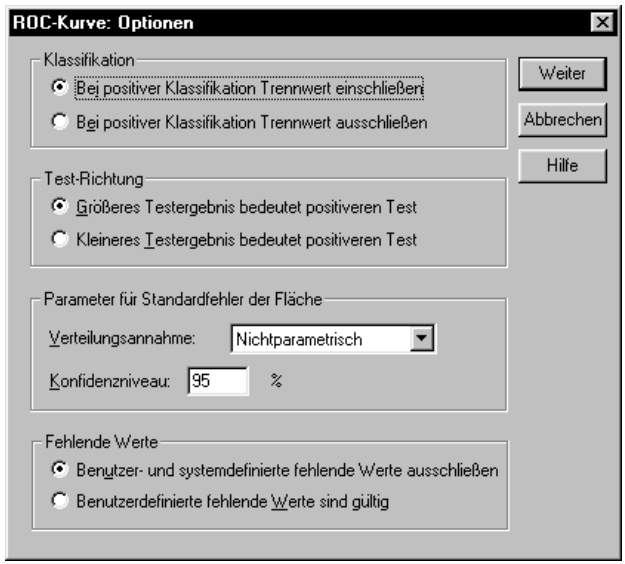
  Grafiken
  ROC-Kurve...

Abbildung 42-2
Dialogfeld "ROC-Kurve"

- Wählen Sie mindestens eine Wahrscheinlichkeitsvariable für den Test aus.
- Wählen Sie eine Zustandsvariable aus.
- Legen Sie den *positiven* Wert für die Zustandsvariable fest.

ROC-Kurve: Optionen

Abbildung 42-3
Dialogfeld "ROC-Kurve: Optionen"



Sie können eine der folgenden Optionen für die ROC-Analyse auswählen:

Klassifikation. Hiermit können Sie festlegen, ob der Trennwert bei einer *positiven* Klassifikation einbezogen oder ausgeschlossen werden soll. Diese Einstellung hat gegenwärtig keine Auswirkungen auf die Ausgabe.

Test-Richtung. Hiermit geben Sie die Richtung der Skala bezogen auf die *positive* Kategorie an.

Parameter für Standardfehler der Fläche. Hiermit geben Sie die Methode an, mit welcher der Standardfehler der Fläche unter der Kurve geschätzt wird. Es stehen eine nichtparametrische und eine binegative exponentielle Methode zur Verfügung. Sie können hier außerdem das Niveau des Konfidenzintervalls festlegen. Es sind Werte zwischen 50,1% und 99,9% möglich.

Fehlende Werte. Hier können Sie festlegen, wie fehlende Werte behandelt werden.

Kapitel
43

Extras

In diesem Kapitel werden die Funktionen aus dem Menü “Extras” und das Umsortieren von Listen mit Zielvariablen unter der Verwendung des Windows-Systemmenüs behandelt.

Variablenbeschreibungen

Im Dialogfeld “Variablen” werden die folgenden Informationen zur Definition der aktuell ausgewählten Variablen angezeigt:

- Datenformat
- Variablenlabel
- Benutzerdefinierte fehlende Werte
- Wertelabel

Abbildung 43-1
Dialogfeld “Variablen”

Gehe zu. Hiermit können Sie im Fenster des Daten-Editors zur ausgewählten Variable wechseln.

Einfügen. Hiermit können Sie die ausgewählten Variablen im Haupt-Syntax-Fenster an der Cursorposition einfügen.

Verwenden Sie die Variablenansicht des Daten-Editors zum Ändern der Definitionen von Variablen.

So rufen Sie Variablenbeschreibungen auf

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Extras
  Variablen...

- Wählen Sie die Variable aus, für die Sie Informationen zur Definition aufrufen möchten.

Datendateikommentare

Sie können beschreibende Kommentare in die Datendateien aufnehmen. Bei Datendateien im SPSS-Format werden diese Kommentare zusammen mit den Datendateien gespeichert.

So können Sie Kommentare zu Datendateien hinzufügen, bearbeiten, löschen und anzeigen

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Extras
  Datendateikommentare...

- Um die Kommentare im Viewer anzuzeigen, wählen Sie die Option Kommentare in Ausgabe anzeigen.

Kommentare können beliebig lang sein, sind jedoch auf 80 Byte (bei Single-Byte-Sprachen entspricht dies normalerweise 80 Zeichen) pro Zeile begrenzt; die Zeilen brechen automatisch nach 80 Zeichen um. Kommentare werden in derselben Schriftart angezeigt wie die Textausgabe, um genau widerzuspiegeln, wie sie bei der Anzeige im Viewer dargestellt werden.

Immer, wenn Sie Kommentare hinzufügen oder bearbeiten, wird automatisch ein Datumsstempel (das aktuelle Datum in Klammern) an das Ende der Kommentarliste angehängt. Dies kann zu Unklarheiten hinsichtlich des den Kommentaren zuzuordnenden Datums führen, wenn Sie einen bestehenden Kommentar bearbeiten oder einen neuen Kommentar zwischen bestehenden Kommentaren einfügen.

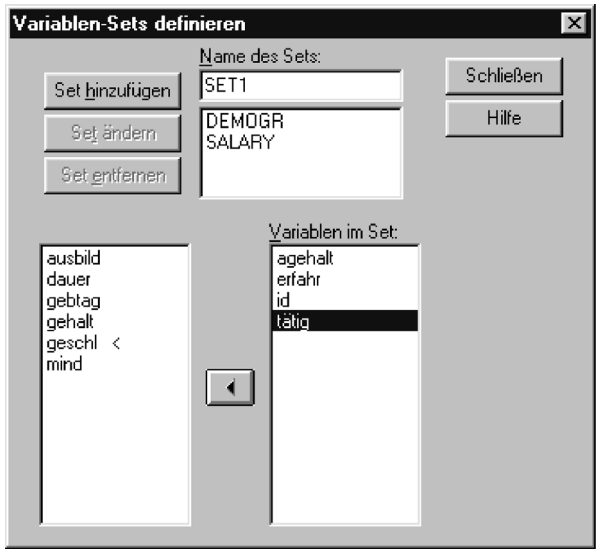
Variablen-Sets

Durch Definieren und Verwenden von Variablen-Sets können Sie einschränken, welche Variablen in der Liste der Quellvariablen in Dialogfeldern angezeigt werden. Dies ist insbesondere bei Datendateien mit einer großen Anzahl an Variablen nützlich. Kleine Variablen-Sets erleichtern das Auffinden und Auswählen von Variablen für die Analyse und können darüber hinaus die Geschwindigkeit von SPSS erhöhen. Wenn Ihre Datendatei eine große Anzahl von Variablen enthält und das Öffnen von Dialogfeldern lange dauert, kann das Einschränken der Listen der Quellvariablen in Dialogfeldern auf kleinere Teil-Sets von Variablen das Öffnen von Dialogfeldern beschleunigen.

Variablen-Sets definieren

Mit “Variablen-Sets definieren” können Sie Teilmengen von Variablen zum Anzeigen in den Listen der Quellvariablen in Dialogfeldern erstellen.

Abbildung 43-2
Dialogfeld "Variablen-Sets definieren"



Name des Sets. Die Namen von Sets können eine Länge von bis zu 12 Zeichen aufweisen. Alle Zeichen, einschließlich Leerzeichen, können verwendet werden. Zwischen Groß- und Kleinbuchstaben wird bei Namen für Sets nicht unterschieden.

Variablen im Set. Ein Set kann eine beliebige Kombination aus numerischen Variablen sowie kurzen und langen String-Variablen enthalten. Die Reihenfolge der Variablen im Set hat keine Auswirkung auf die Anzeigereihenfolge der Variablen in der Liste der Quellvariablen in Dialogfeldern. Eine Variable kann in verschiedenen Sets enthalten sein.

So definieren Sie Variablen-Sets:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Extras
  Sets definieren…

▶ Wählen Sie die Variablen aus, die Sie in das Set aufnehmen möchten.

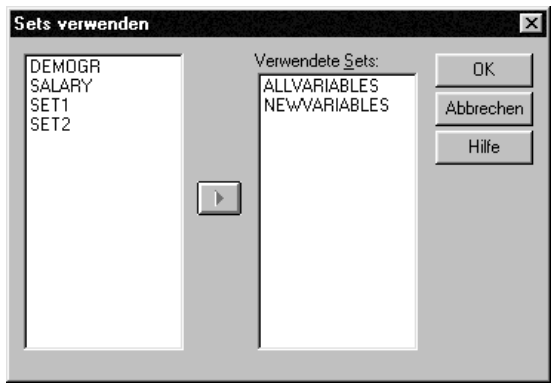
▶ Geben Sie einen Namen für das Set ein. Der Name darf aus bis zu zwölf Zeichen bestehen.

▶ Klicken Sie auf Set hinzufügen.

Sets verwenden

Mit “Sets verwenden” können Sie festlegen, daß in den Listen der Quellvariablen in Dialogfeldern nur die Variablen aus den ausgewählten Sets angezeigt werden.

Abbildung 43-3
Dialogfeld “Sets verwenden”



Verwendete Sets. Hier werden die Sets angezeigt, auf denen die Listen der Quellvariablen in Dialogfeldern beruhen. In der Liste der Quellvariablen werden die Variablen alphabetisch geordnet, oder die Reihenfolge der Variablen aus der Datei wird beibehalten. Die Reihenfolge der Sets und die Reihenfolge der Variablen in einem Set haben keine Auswirkung auf die Reihenfolge der Variablen in der Liste der Quellvariablen. Als Standardeinstellung werden die folgenden zwei systemdefinierten Sets verwendet:

ALLVARIABLES. Dieses Set enthält alle Variablen in der Datendatei sowie die neuen Variablen, die in einer Sitzung erstellt werden.

NEWVARIABLES. Dieses Set enthält nur die in einer Sitzung erstellten Variablen.

Sie können diese Sets aus der Liste entfernen und andere Sets auswählen. Die Liste muß jedoch mindestens ein Set enthalten. Solange Sie das Set *ALLVARIABLES* nicht aus der Liste “Verwendete Sets” entfernen, wird die Anzahl der Variablen durch kein anderes Set eingeschränkt.

So beschränken Sie die Listen der Quellvariablen in Dialogfeldern auf die definierten Variablen-Sets:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

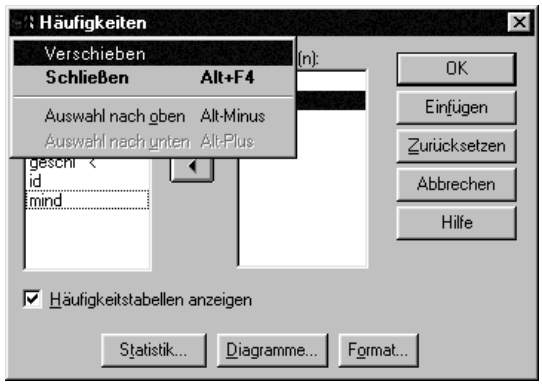
  Extras
  Sets verwenden…

- Wählen Sie die Variablen-Sets, welche die Variablen enthalten, die in den Listen der Quellvariablen in Dialogfeldern angezeigt werden sollen.

Umsortieren von Listen mit Zielvariablen

Die Variablen in den Listen der Zielvariablen in Dialogfeldern werden in der Reihenfolge angezeigt, in der diese aus der Liste der Quellvariablen ausgewählt wurden. Wenn Sie die Reihenfolge der Variablen in der Liste der Zielvariablen ändern möchten, aber nicht die Auswahl aller Variablen aufheben und dann alle Variablen erneut und in anderer Reihenfolge auswählen möchten, können Sie die Variablen mit dem Systemmenü in der oberen linken Ecke des Dialogfelds in der Liste der Zielvariablen nach oben oder nach unten verschieben. Klicken Sie hierzu auf das linke Ende der Titelleiste des Dialogfelds.

Abbildung 43-4
Windows-Systemmenü und das Umsortieren der Liste der Zielvariablen



Auswahl nach oben. Hiermit verschieben Sie die ausgewählten Variablen in der Liste der Zielvariablen um eine Position nach oben.

Auswahl nach unten. Hiermit verschieben Sie die ausgewählten Variablen in der Liste der Zielvariablen um eine Position nach unten.

Sie können mehrere Variablen auf einmal verschieben, wenn die Variablen in der Liste direkt aufeinanderfolgen. Sie können keine Gruppen von Variablen verschieben, wenn die Variablen in der Liste nicht direkt aufeinander folgen.

Kapitel
44

Optionen

Im Dialogfeld “Optionen” können Sie eine Vielzahl von Einstellungen ändern, darunter:

- Das Sitzungs-Journal, in dem alle in einer Sitzung verwendeten Befehle aufgezeichnet werden
- Die Reihenfolge der Anzeige von Variablen in den Quellisten von Dialogfeldern
- Angezeigte und ausgeblendete Objekte in neu ausgegebenen Ergebnissen
- Die Tabellenvorlage für neue Pivot-Tabellen und die Diagrammvorlage für neue interaktive Diagramme
- Währungsformate
- Autoskript-Dateien und Autoskript-Funktionen zur benutzerdefinierten Anpassung der Ausgabe

So ändern Sie die Einstellungen für die Optionen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Bearbeiten
  Optionen...

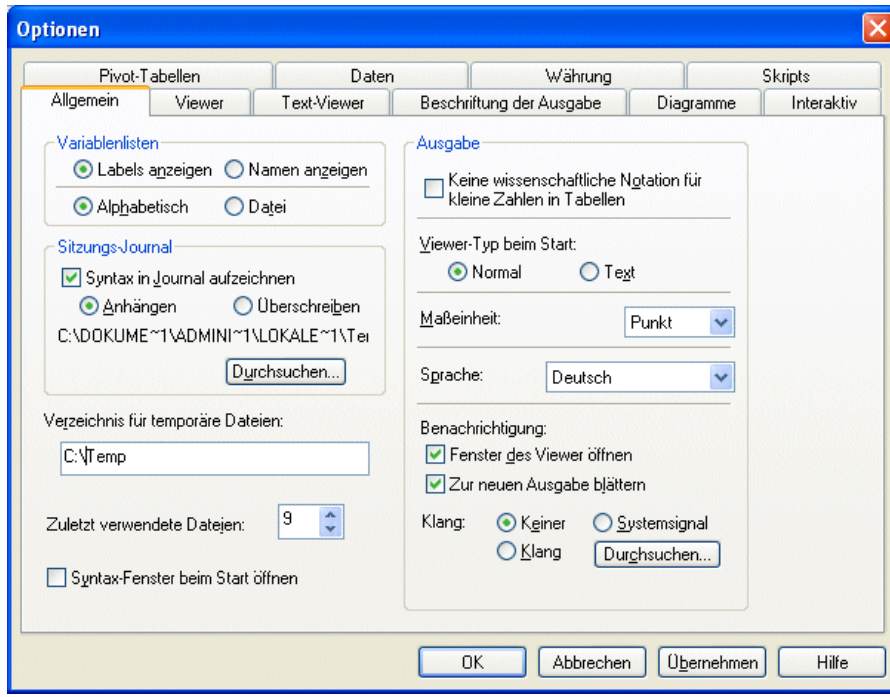
▶ Klicken Sie zum Ändern der Einstellungen auf die entsprechende Registerkarte.

▶ Ändern Sie die Einstellungen.

▶ Klicken Sie auf OK oder Zuweisen.

Optionen: Allgemein

Abbildung 44-1
Dialogfeld "Optionen", Registerkarte "Allgemein"



Variablenlisten. Hiermit legen Sie fest, wie Variablen in den Listen der Dialogfelder angezeigt werden sollen. Sie können Variablennamen oder Variablenlabels anzeigen lassen. Die Namen oder Labels von Variablen können in alphabetischer oder in der Reihenfolge sortiert werden, in welcher die Variablen in der Datei (und im Fenster des Daten-Editors) vorkommen. Die Reihenfolge der Anzeige wirkt sich nur auf Listen von Quellvariablen aus. Listen von Zielvariablen geben immer die Reihenfolge wieder, in welcher die Variablen ausgewählt wurden.

Sitzungs-Journal. Die Journaldatei aller in einer Sitzung ausgeführten Befehle. Dazu gehören Befehle, die in Syntax-Fenstern eingegeben und von dort ausgeführt wurden, und aus Dialogfeldern aufgerufene Befehle. Sie können die Journaldatei bearbeiten und die Befehle erneut in anderen Sitzungen verwenden. Sie können das Führen der Journaldatei aktivieren und deaktivieren, etwas an die Journaldatei anhängen oder

die Datei überschreiben sowie Namen und Speicherort der Journaldatei auswählen. Sie können die Befehlssyntax aus der Journaldatei kopieren und in einer Syntaxdatei speichern, um sie im SPSS-Produktionsmodus zu verwenden.

Verzeichnis für temporäre Dateien. Hiermit legen Sie den Speicherort für die temporären Dateien fest, die in einer Sitzung erstellt werden. Der Speicherort für temporäre Datendateien im Modus für verteilte Analysen (verfügbar mit der Server-Version) wird hierdurch nicht beeinflußt. Im Modus für verteilte Analysen wird der Speicherort für temporäre Dateien durch die Umgebungsvariable *SPSSTMPDIR* festgelegt. Diese Variable kann nur auf dem Computer gesetzt werden, auf dem die Server-Version der Software ausgeführt wird. Wenn Sie den Speicherort für temporäre Dateien ändern möchten, wenden Sie sich an Ihren Systemadministrator.

Zuletzt verwendete Dateien. Hiermit legen Sie die Anzahl der im Menü "Datei" angezeigten zuletzt verwendeten Dateien fest.

Syntax-Fenster beim Starten öffnen. Syntax-Fenster sind Fenster für Textdateien zum Eingeben, Bearbeiten und Ausführen von Befehlen. Wenn Sie häufig mit der Befehlssyntax arbeiten, können Sie mit dieser Option zu Beginn jeder SPSS-Sitzung automatisch ein Syntax-Fenster öffnen. Dies bietet sich in erster Linie für erfahrene Anwender an, die lieber mit der Befehlssprache als mit Dialogfeldern arbeiten. (In der Studentenversion nicht verfügbar.)

Keine wissenschaftliche Notation für kleine Zahlen in Tabellen. Hiermit wird die wissenschaftliche Notation bei kleinen Dezimalwerten in der Ausgabe unterdrückt. Sehr kleine Dezimalwerte werden als 0 (oder 0,000) angezeigt.

Viewer-Typ beim Start. Hiermit legen Sie den Viewer-Typ und das Ausgabeformat fest. Der Viewer erzeugt interaktive Pivot-Tabellen und interaktive Diagramme. Der Text-Viewer wandelt Pivot-Tabellen in Textausgabe und Diagramme in Metadateien um.

Maßeinheit. Die Maßeinheit (Punkt, Zoll oder Zentimeter), in der Druckparameter, wie zum Beispiel die Zellenränder von Pivot-Tabellen, Zellenbreiten und Abstand zwischen Tabellen, angegeben werden.

Sprache. Hiermit können Sie die in der Ausgabe verwendete Sprache festlegen. Dies gilt jedoch nicht für einfache Textausgaben, interaktive Grafiken und Karten (verfügbar mit dem Erweiterungsmodul "Maps"). Es hängt von den installierten Sprachdateien ab, welche Sprachen verfügbar sind.

Anmerkung: Benutzerdefinierte Skripts, die sich auf sprachspezifische Zeichenfolgen oder Textpassagen in der Ausgabe stützen, werden möglicherweise nicht ordnungsgemäß ausgeführt, wenn Sie die Ausgabesprache ändern. Für weitere Informationen siehe “Optionen: Skripts” auf S. 684.

Benachrichtigung. Hiermit können Sie festlegen, wie das Programm Sie über den Abschluß einer Prozedur und die Anzeige der Ergebnisse im Viewer benachrichtigen soll.

Optionen: Viewer

Die Optionen für die Anzeige der Ausgaben wirken sich nur auf die nach dem Ändern der Einstellungen erzeugten Ausgaben aus. Änderungen der Einstellungen wirken sich nicht auf bereits angezeigte Ausgaben aus.

Abbildung 44-2
Dialogfeld “Optionen”, Registerkarte “Viewer”

Anfänglicher Ausgabestatus. Hiermit legen Sie fest, welche Objekte bei jeder Ausführung einer Prozedur automatisch angezeigt oder ausgeblendet werden und wie die Objekte ausgerichtet sind. Sie können festlegen, daß folgende Objekte angezeigt bzw. nicht angezeigt werden: Log, Warnungen, Anmerkungen, Titel, Pivot-Tabellen, Diagramme und Textausgabe (nicht in Form von Pivot-Tabellen angezeigte Ausgabe). Außerdem können Sie die Anzeige von Befehlen im Log aktivieren oder deaktivieren. Sie können Befehlssyntax aus dem Log kopieren und in einer Syntaxdatei speichern, um diese im SPSS-Produktionsmodus zu verwenden.

Anmerkung: Im Viewer werden alle Ausgabeobjekte linksbündig angezeigt. Die Einstellungen für die Ausrichtung wirken sich nur auf gedruckte Ausgaben aus. Zentrierte und rechtsbündig ausgerichtete Objekte werden durch kleine Symbole an der linken oberen Ecke des Objekts gekennzeichnet.

Schriftart für Titel. Hiermit können Sie die Schriftart, Schriftgröße und die Farbe für die Titel neuer Ausgaben festlegen.

Schriftart für Seitentitel. Hiermit legen Sie die Schriftart, Schriftgröße und die Farbe für neue Seitentitel fest, die mit dem Befehl Neuer Seitentitel im Menü "Einfügen" erstellt wurden, und für Seitentitel, die durch die Syntaxbefehl `TITLE` und `SUBTITLE` erzeugt wurden.

Seitengröße für Textausgabe. Hiermit können Sie für die Textausgabe Seitenbreite und Seitenlänge einstellen. Die Seitenbreite wird in Anzahl an Zeichen, die Seitenlänge in Anzahl an Zeilen angegeben. Bei einigen Prozeduren werden einige Statistiken nur im Querformat angezeigt.

Schriftart für Textausgabe. Dies ist die in der Textausgabe verwendete Schriftart. Für die Textausgabe sollte ein Zeichensatz mit festem Abstand verwendet werden. Bei Einsatz von Proportionalschriften werden Tabellenausgaben fehlerhaft ausgerichtet. Die Schriftart der Textausgabe wird außerdem im Text-Assistenten verwendet, um den Inhalt von Dateien anzuzeigen. Auch hierbei wird der Text falsch ausgerichtet, wenn es sich um eine nicht proportionale Schriftart handelt.

Optionen: Text-Viewer

Die Optionen für die Anzeige der Ausgaben im Text-Viewer wirken sich nur auf die Ausgaben aus, die nach dem Ändern der Einstellungen berechnet werden. Änderungen der Einstellungen wirken sich nicht auf die im Text-Viewer bereits angezeigte Ausgabe aus.

Abbildung 44-3
Dialogfeld "Optionen", Registerkarte "Text-Viewer"

Ausgabeelemente anzeigen. Hiermit können Sie festlegen, welche Objekte beim Ausführen einer Prozedur automatisch angezeigt werden. Sie können festlegen, daß folgende Objekte angezeigt bzw. nicht angezeigt werden: Log, Warnungen, Anmerkungen, Titel, Tabellenausgabe (in Textausgabe umgewandelte Pivot-Tabellen), Diagramme und Textausgabe (durch Leerzeichen getrennte Ausgabe). Außerdem können Sie die Anzeige von Befehlen im Log aktivieren oder deaktivieren. Sie können Befehlssyntax aus dem Log kopieren und in einer Syntaxdatei speichern, um diese im SPSS-Produktionsmodus zu verwenden.

Seitenumbruch zwischen. Hiermit können Sie einen Seitenumbruch zwischen Ausgaben von verschiedenen Prozeduren und/oder zwischen einzelnen Ausgabeobjekten einfügen.

Schriftart. Die für die neue Ausgabe verwendete Schriftart. Hier stehen nur Zeichensätze mit festem Abstand zur Verfügung, da durch Leerzeichen getrennte Textausgabe mit Proportionalschriften fehlerhaft ausgerichtet wird.

Tabellenausgabe. Hiermit können Sie Einstellungen für die Ausgabe von Pivot-Tabellen vornehmen, die in tabellarische Textausgabe umgewandelt wurde. Die Angaben für die Spaltenbreite und das Spaltentrennzeichen sind nur verfügbar, wenn Sie Leerzeichen als Spaltentrennzeichen auswählen. Bei durch Leerzeichen getrennter Tabellenausgabe werden standardmäßig alle Zeilenumbrüche entfernt, und die Spaltenbreite richtet sich nach der Breite der längsten Beschriftung oder des längsten Werts in der Spalte. Wenn Sie die Spaltenbreite beschränken und lange Beschriftungen umbrechen möchten, geben Sie eine Anzahl von Zeichen für die Spaltenbreite an.

Anmerkung: Durch Tabulatoren getrennte Tabellenausgabe wird im Text-Viewer fehlerhaft ausgerichtet. Dieses Format ist sinnvoll, wenn Sie Ergebnisse durch Kopieren und Einfügen in ein Textverarbeitungsprogramm übernehmen möchten, in dem Sie alle Schriftarten (nicht nur solche mit festem Zeichenabstand) verwenden und die Ausgabe mit Tabulatoren ausrichten können.

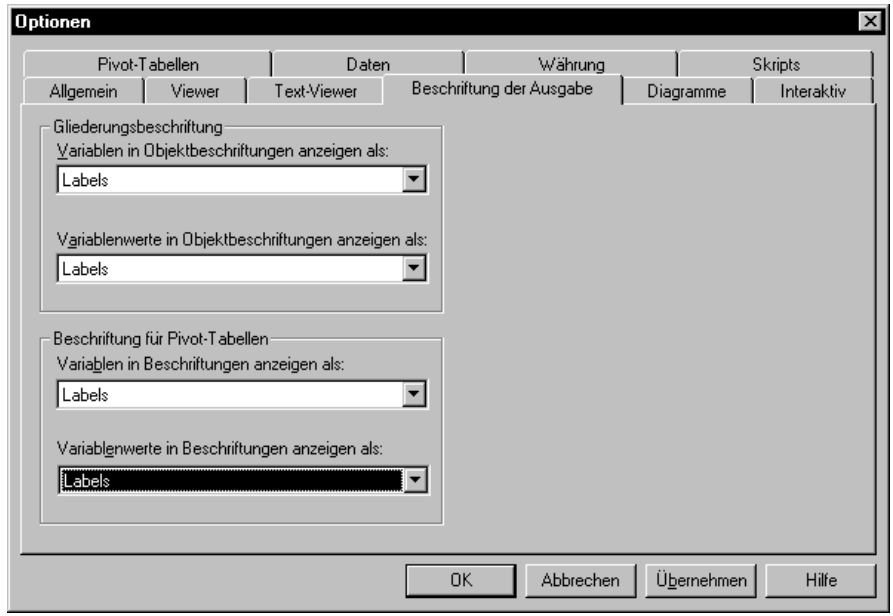
Textausgabe. Bei Textausgaben, bei denen es sich nicht um umgewandelte Pivot-Tabellen handelt, wird hiermit die Seitenbreite (in Anzahl von Zeichen) und die Seitenlänge (in Anzahl von Zeilen) festgelegt. Bei einigen Prozeduren werden einige Statistiken nur im Querformat angezeigt.

Optionen: Beschriftung der Ausgabe

Mit den Optionen für die Beschriftung der Ausgabe können Sie Einstellungen für die Anzeige der Variablen und Datenwerte in der Gliederung und in Pivot-Tabellen vornehmen. Sie können Variablennamen, definierte Variablenlabels und Datenwerte, definierte Wertelabels oder eine Kombination dieser Angaben anzeigen lassen.

Aussagekräftige Variablen- und Wertelabels (Variablenansicht im Daten-Editor, Spalten *Variablenlabel* und *Wertelabels*) können die Interpretation der Ergebnisse häufig erleichtern. In manchen Tabellen können lange Labels jedoch eher stören.

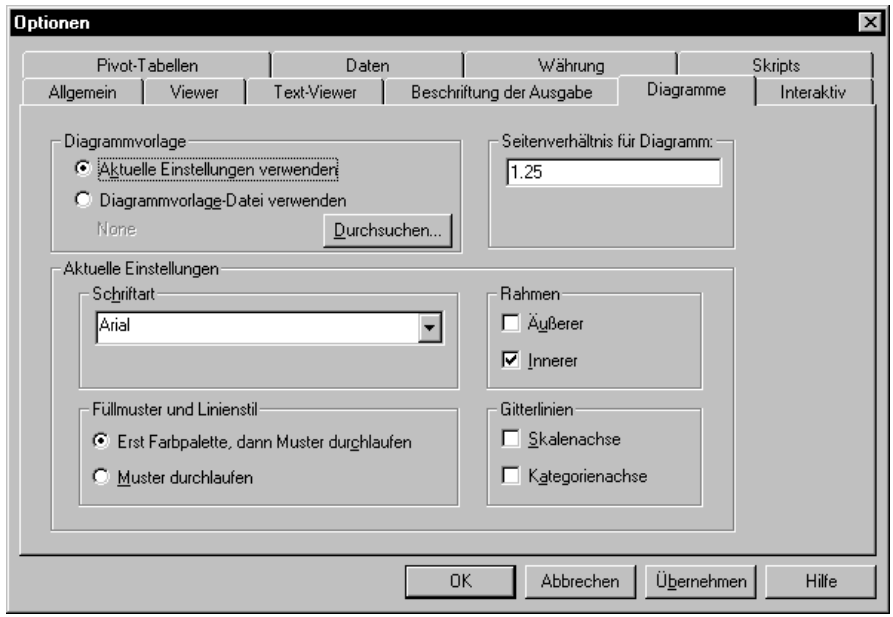
Abbildung 44-4
Dialogfeld "Optionen", Registerkarte "Beschriftung der Ausgabe"



Die Optionen für die Beschriftung der Ausgabe wirken sich nur auf neue Ausgaben aus, die nach der Änderung der Einstellungen berechnet werden. Änderungen der Einstellungen wirken sich nicht auf bereits angezeigte Ausgaben aus. Diese Einstellungen betreffen nur die Ausgabe von Pivot-Tabellen. Sie wirken sich nicht auf die Textausgabe aus.

Diagrammoptionen

Abbildung 44-5
Dialogfeld "Optionen", Registerkarte "Diagramme"



Diagrammvorlage. Sie können für neue Diagramme entweder die hier ausgewählten Einstellungen oder die Einstellungen aus einer Diagrammvorlagendatei auswählen. Klicken Sie auf Durchsuchen, um eine Diagrammvorlagendatei auszuwählen. Sie erstellen eine Diagrammvorlagendatei, indem Sie eine Datei mit den von Ihnen gewünschten Merkmalen erstellen und im Menü "Datei" den Befehl Diagrammvorlage speichern auswählen.

Seitenverhältnis für Diagramm. Dies ist das Verhältnis von Breite zu Höhe des äußeren Rahmens neuer Diagramme. Sie können ein Verhältnis von Breite zu Höhe von 0,1 bis 10,0 festlegen. Werte unter 1 ergeben Diagramme im Hochformat. Werte größer als 1 ergeben Diagramme im Querformat. Ein Wert von 1 ergibt quadratische Diagramme. Bei bereits erstellten Diagrammen kann das Seitenverhältnis nicht mehr geändert werden.

JVM beim Programmstart starten. Für Diagrammfunktionen ist die Java Virtual Machine (JVM) erforderlich. In der Standardeinstellung wird die JVM gestartet, wenn Sie eine SPSS-Sitzung beginnen. Wenn Sie diese Option deaktiveren, startet SPSS möglicherweise schneller. Es wird jedoch eine kleine Verzögerung auftreten, um die JVM zu starten, wenn Sie das erste Mal in der Sitzung die Diagrammerstellung verwenden oder ein Diagramm erstellen.

Schriftart. Diese Schriftart wird für den gesamten Text in neuen Diagrammen verwendet.

Bevorzugte Stilauswahlmethode. Hiermit geben Sie die Anfangswerte für Farben und/oder Muster für neue Diagramme ein. Bei Nur Farben durchlaufen werden zum Unterscheiden verschiedener Diagrammelemente keine Muster, sondern nur Farben verwendet. Bei Nur Muster durchlaufen werden zum Unterscheiden verschiedener Diagrammelemente keine Farben, sondern nur Linienstile, Markierungssymbole und Füllmuster verwendet.

Rahmen. Hiermit können Sie festlegen, ob neue Diagramme mit inneren bzw. äußeren Rahmen erstellt werden sollen.

Gitterlinien. Hiermit können Sie festlegen, ob neue Diagramme mit Gitterlinien für die Skalen- und Kategorienachse angezeigt werden sollen.

Stilauswahlmethoden. Dient zur benutzerdefinierten Anpassung der Farben, Linienstile, Markierungssymbole und Füllmuster für neue Diagramme. Sie können die Anordnung der Farben und Muster ändern, die beim Erstellen eines neuen Diagramms verwendet werden.

Anmerkung: Diese Einstellungen wirken sich nicht auf interaktive Diagramme aus (Menü “Grafiken”, Untermenü “Interaktiv”).

Farben der Datenelemente

Geben Sie die Reihenfolge an, in der die Farben im neuen Diagramm für die Datenelemente (z. B. Balken und Markierungen) verwendet werden sollen. Farben werden immer dann verwendet, wenn Sie eine Auswahl treffen, zu der im Hauptdialogfeld “Diagrammoptionen” in der Gruppe “Bevorzugte Stilauswahlmethode” die Option *Farbe* gehört.

Wenn Sie beispielsweise ein gruppiertes Balkendiagramm mit zwei Gruppen erstellen und im Hauptdialogfeld "Diagrammoptionen" die Option Erst Farbpalette, dann Muster durchlaufen auswählen, werden die ersten beiden Farben in der Liste der gruppierten Diagramme im neuen Diagramm als Balkenfarben verwendet.

So ändern Sie die Reihenfolge, in der die Farben verwendet werden

- Wählen Sie die Option Einfache Diagramme, und wählen Sie dann eine Farbe aus, die für Diagramme ohne Kategorien verwendet werden soll.

- Wählen Sie die Option Gruppierte Diagramme, um die Farbauswahlmethode für Diagramme mit Kategorien zu ändern. Wenn Sie die Farbe einer Kategorie ändern möchten, wählen Sie die Kategorie und anschließend eine Farbe für diese Kategorie aus der Farbpalette aus.

Die folgenden Optionen sind verfügbar:

- Sie können vor der jeweils ausgewählten Kategorie eine neue Kategorie einfügen.
- Ausgewählte Kategorien können verschoben werden.
- Ausgewählte Kategorien können entfernt werden.
- Die Sequenz kann auf den Standardwert zurückgesetzt werden.
- Sie können eine Farbe bearbeiten, indem Sie ihre Quelle auswählen und auf Bearbeiten klicken.

Linien von Datenelementen

Geben Sie die Reihenfolge an, in der die verschiedenen Stile für linienförmige Datenelemente in Ihrem neuen Diagramm verwendet werden sollen. Linienstile werden immer dann verwendet, wenn das Diagramm linienförmige Datenelemente enthält und Sie eine Auswahl treffen, zu der im Hauptdialogfeld "Diagrammoptionen" in der Gruppe "Bevorzugte Stilauswahlmethode" die Option *Muster* gehört.

Wenn Sie beispielsweise ein Liniendiagramm mit zwei Gruppen erstellen und im Hauptdialogfeld "Diagrammoptionen" die Option Nur Muster durchlaufen auswählen, werden die ersten beiden Stile in der Liste der gruppierten Diagramme im neuen Diagramm als Linienmuster verwendet.

So ändern Sie die Reihenfolge, in der die Linienmuster verwendet werden

- Wählen Sie die Option Einfache Diagramme, und wählen Sie dann einen Linienstil aus, der für Liniendiagramme ohne Kategorien verwendet werden soll.
- Wählen Sie die Option Gruppierte Diagramme, um die Musterauswahlmethode für Diagramme mit Kategorien zu ändern. Wenn Sie den Linienstil einer Kategorie ändern möchten, wählen Sie die Kategorie und anschließend einen Linienstil für diese Kategorie aus der Palette aus.

  Die folgenden Optionen sind verfügbar:

  - Sie können vor der jeweils ausgewählten Kategorie eine neue Kategorie einfügen.
  - Ausgewählte Kategorien können verschoben werden.
  - Ausgewählte Kategorien können entfernt werden.
  - Die Sequenz kann auf den Standardwert zurückgesetzt werden.

Markierungen für Datenelemente

Geben Sie die Reihenfolge an, in der die verschiedenen Symbole für Markierungs-Datenelemente in Ihrem neuen Diagramm verwendet werden sollen. Markierungsstile werden immer dann verwendet, wenn das Diagramm Markierungs-Datenelemente enthält und Sie eine Auswahl treffen, zu der im Hauptdialogfeld "Diagrammoptionen" in der Gruppe "Bevorzugte Stilauswahlmethode" die Option *Muster* gehört.

Wenn Sie beispielsweise ein Streudiagramm mit zwei Gruppen erstellen und im Hauptdialogfeld "Diagrammoptionen" die Option Nur Muster durchlaufen auswählen, werden die ersten beiden Symbole in der Liste der gruppierten Diagramme im neuen Diagramm als Markierungen verwendet.

So ändern Sie die Reihenfolge, in der die Markierungsstile verwendet werden

- Wählen Sie die Option Einfache Diagramme, und wählen Sie dann ein Symbol aus, das für Diagramme ohne Kategorien verwendet werden soll.
- Wählen Sie die Option Gruppierte Diagramme, um die Musterauswahlmethode für Diagramme mit Kategorien zu ändern. Wenn Sie das Markierungssymbol einer Kategorie ändern möchten, wählen Sie die Kategorie und anschließend ein Symbol für diese Kategorie aus der Palette aus.

Die folgenden Optionen sind verfügbar:

- Sie können vor der jeweils ausgewählten Kategorie eine neue Kategorie einfügen.
- Ausgewählte Kategorien können verschoben werden.
- Ausgewählte Kategorien können entfernt werden.
- Die Sequenz kann auf den Standardwert zurückgesetzt werden.

Füllmuster für Datenelemente

Geben Sie die Reihenfolge an, in der die verschiedenen Füllstile für Balken- und Flächen-Datenelemente in Ihrem neuen Diagramm verwendet werden sollen. Füllstile werden immer dann verwendet, wenn das Diagramm Balken- oder Flächen-Datenelemente enthält und Sie eine Auswahl treffen, zu der im Hauptdialogfeld "Diagrammoptionen" in der Gruppe "Bevorzugte Stilauswahlmethode" die Option *Muster* gehört.

Wenn Sie beispielsweise ein gruppiertes Balkendiagramm mit zwei Gruppen erstellen und im Hauptdialogfeld "Diagrammoptionen" die Option Nur Muster durchlaufen auswählen, werden die ersten beiden Stile in der Liste der gruppierten Diagramme im neuen Diagramm als Füllmuster für die Balken verwendet.

So ändern Sie die Reihenfolge, in der die Füllstile verwendet werden

▶ Wählen Sie die Option Einfache Diagramme, und wählen Sie dann ein Füllmuster aus, das für Diagramme ohne Kategorien verwendet werden soll.

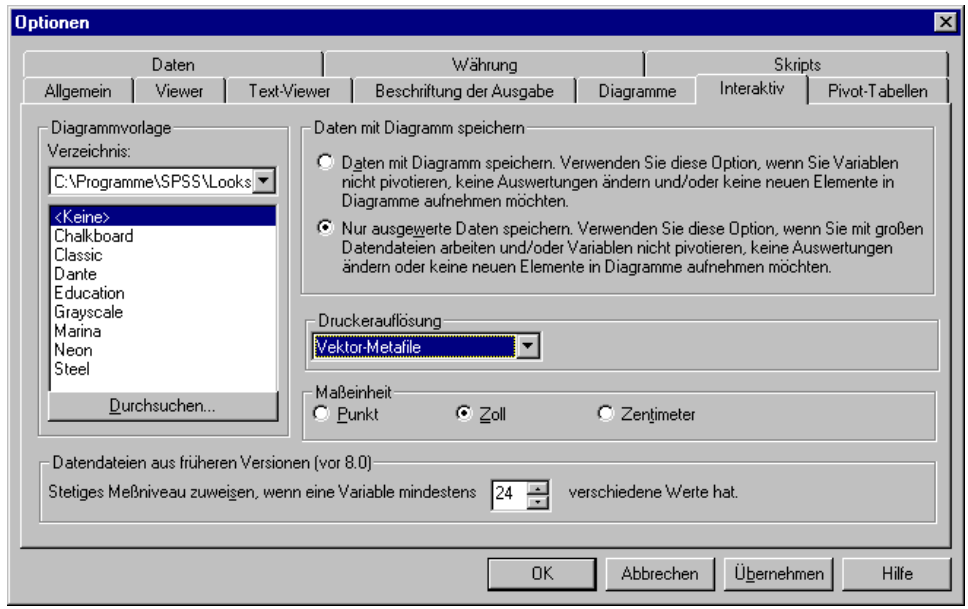
▶ Wählen Sie die Option Gruppierte Diagramme, um die Musterauswahlmethode für Diagramme mit Kategorien zu ändern. Wenn Sie das Füllmuster einer Kategorie ändern möchten, wählen Sie die Kategorie und anschließend ein Füllmuster für diese Kategorie aus der Palette aus.

Die folgenden Optionen sind verfügbar:

- Sie können vor der jeweils ausgewählten Kategorie eine neue Kategorie einfügen.
- Ausgewählte Kategorien können verschoben werden.
- Ausgewählte Kategorien können entfernt werden.
- Die Sequenz kann auf den Standardwert zurückgesetzt werden.

Optionen: Interaktiv

Abbildung 44-6
Dialogfeld "Optionen", Registerkarte "Interaktiv"



Für interaktive Diagramme (Menü "Grafiken", Untermenü "Interaktiv") sind folgende Optionen verfügbar:

Diagrammvorlage. Wählen Sie eine Diagrammvorlage aus der Dateiliste aus, und klicken Sie auf OK oder Übernehmen. In der Standardeinstellung werden in der Liste die im Unterverzeichnis *Looks* des Programmverzeichnisses gespeicherten Diagrammvorlagen angezeigt. Sie können eine der mit dem Programm gelieferten Diagrammvorlagen verwenden oder im Dialogfeld "Diagrammvorlage" Ihre eigene Diagrammvorlage erstellen. Wählen Sie dazu in einem aktivierten Diagramm im Menü "Format" den Befehl Diagrammvorlagen.

- **Verzeichnis.** Hiermit können Sie ein Verzeichnis für Diagrammvorlagen auswählen. Verwenden Sie die Schaltfläche Durchsuchen, um weitere Verzeichnisse in die Liste aufzunehmen.
- **Durchsuchen.** Hiermit können Sie eine Tabellenvorlage aus einem anderen Verzeichnis auswählen.

Daten mit Diagramm speichern. Hiermit können Sie festlegen, welche Informationen mit interaktiven Diagrammen gespeichert werden, wenn die Diagramme nicht mehr an die Datendatei angehängt sind, mit der sie erstellt wurden (zum Beispiel, wenn Sie eine in einer früheren Sitzung gespeicherte Viewer-Datei öffnen). Wenn Sie die Daten mit dem Diagramm speichern, sind die meisten interaktiven Funktionen verfügbar, die auch für Diagramme verfügbar sind, die an die Datendatei angehängt sind, mit der sie erstellt wurden. Sie können allerdings keine Variablen hinzufügen, die im ursprünglichen Diagramm nicht enthalten waren. Dies kann die Viewer-Dateien jedoch erheblich vergrößern, besonders bei großen Datendateien.

Druckerauflösung. Hiermit können Sie die Druckerauflösung interaktiver Diagramme festlegen. In den meisten Fällen ergibt die Option Vektor-Metafile die schnellste Druckausgabe und die besten Ergebnisse. Bei Bitmaps werden Diagramme mit niedriger Auflösung schneller gedruckt, und Diagramme mit hoher Auflösung sehen besser aus.

Maßeinheit. Die verwendete Maßeinheit (Punkt, Zoll oder Zentimeter) zur Angabe von Attributen wie zum Beispiel der Größe der Datenregion in einem Diagramm.

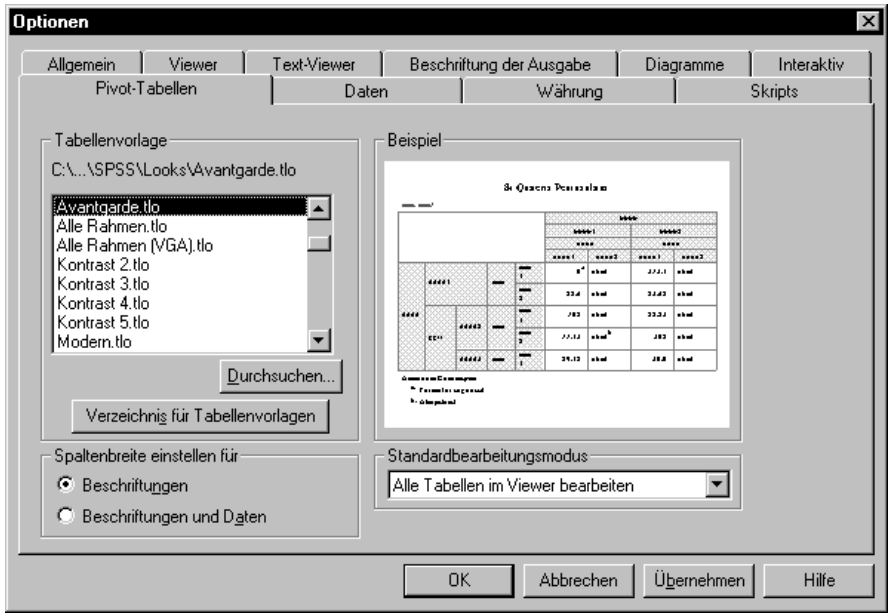
Datendateien aus früheren Versionen (vor 8.0). Für numerische Variablen in Datendateien, die in älteren Versionen von SPSS erstellt wurden, in Daten, die aus externen Dateiformaten eingelesen werden, und in den neuen Variablen, die in einer Sitzung erstellt wurden, können Sie die Mindestanzahl von Datenwerten angeben, anhand derer die Variable als metrische Variable oder kategoriale Variable klassifiziert wird. Variablen mit einer geringeren als der angegebenen Anzahl von eindeutigen Werten werden als nominal klassifiziert.

Anmerkung: Mit Ausnahme des Messniveaus wirken sich diese Einstellungen nur auf interaktive Diagramme aus (Menü “Grafiken”, Untermenü “Interaktiv”).

Pivottabellenoptionen

Mit den Optionen für Pivot-Tabellen können Sie die Standard-Tabellenvorlage einstellen, die für neue Pivot-Tabellen verwendet werden soll. Mit den Tabellenvorlagen können Sie eine Reihe von Parametern für Pivot-Tabellen einstellen, darunter die Anzeige und Breite von Gitterlinien, Schriftart, Schriftgröße und -farbe sowie Hintergrundfarben.

Abbildung 44-7
Dialogfeld "Optionen", Registerkarte "Pivot-Tabellen"



Tabellenvorlage. Wählen Sie eine Tabellenvorlage aus der Dateiliste aus, und klicken Sie auf OK oder Übernehmen. In der Standardeinstellung werden in der Liste die im Unterverzeichnis *Looks* des Programmverzeichnisses gespeicherten Tabellenvorlagen angezeigt. Sie können eine der mit dem Programm gelieferten Tabellenvorlagen verwenden oder im Dialogfeld "Tabellenvorlagen" Ihre eigene Tabellenvorlage erstellen. Wählen Sie dazu aus dem Menü "Format" den Befehl Tabellenvorlagen.

- **Durchsuchen.** Hiermit können Sie eine Tabellenvorlage aus einem anderen Verzeichnis auswählen.
- **Verzeichnis für Tabellenvorlagen.** Hiermit können Sie ein anderes Standardverzeichnis für Tabellenvorlagen angeben.

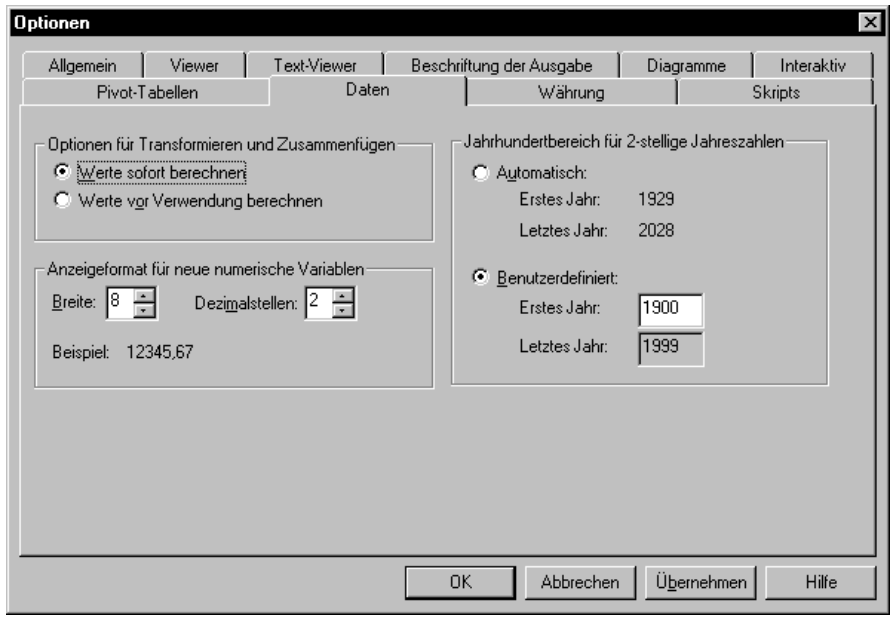
Spaltenbreite einstellen für. Hiermit können Sie die automatische Anpassung der Spaltenbreite in Pivot-Tabellen einstellen.

- **Beschriftungen.** Hiermit passen Sie die Spaltenbreite an die Breite der Spaltenbeschriftung an. Dadurch erhalten Sie kompaktere Tabellen, aber Datenwerte, die breiter als die Beschriftung sind, werden nicht angezeigt. (Sternchen weisen auf Werte hin, die zu breit für die Anzeige sind.)
- **Beschriftungen und Daten.** Hiermit passen Sie die Spaltenbreite an die Spaltenbeschriftung oder den Datenwert an, je nachdem, welcher der beiden länger ist. Dies erzeugt breitere Tabellen, stellt jedoch die Anzeige aller Werte sicher.

Standardbearbeitungsmodus. Hiermit können Sie einstellen, ob Pivot-Tabellen im Viewer-Fenster oder in einem separaten Fenster aktiviert werden. In der Standardeinstellung wird eine Pivot-Tabelle durch Doppelklicken im Viewer-Fenster geöffnet. Sie können Pivot-Tabellen jedoch auch in einem separaten Fenster öffnen oder festlegen, daß kleine Pivot-Tabellen im Viewer-Fenster und Pivot-Tabellen ab einer bestimmten Größe in einem separaten Fenster geöffnet werden.

Optionen: Daten

Abbildung 44-8
Dialogfeld "Optionen", Registerkarte "Daten"



Optionen für Transformieren und Zusammenfügen. Bei jeder Ausführung eines Befehls wird die Datendatei gelesen. Für einige Datentransformationen wie beispielsweise "Berechnen" und "Umkodieren" sowie Dateitransformationen wie beispielsweise "Variablen hinzufügen" und "Fälle hinzufügen" ist kein separater Datendurchlauf erforderlich. Die Ausführung dieser Befehle kann verschoben werden, bis das Programm die Daten liest, um einen anderen Befehl auszuführen, beispielsweise eine statistische Prozedur. Wählen Sie für große Dateien Werte vor Verwendung berechnen aus, um die Ausführung zu verzögern und so Bearbeitungszeit zu sparen.

Anzeigeformat für neue numerische Variablen. Hiermit können Sie die Standardbreite und Anzahl der Dezimalstellen bei der Anzeige neuer numerischer Variablen festlegen. Es gibt kein Standard-Anzeigeformat für neue String-Variablen. Falls ein Wert zu groß für das festgelegte Anzeigeformat ist, werden erst Dezimalstellen gerundet und dann die Werte in wissenschaftliche Notation umgewandelt. Anzeigeformate haben keine Auswirkung auf die internen Datenwerte. So kann der Wert 123456,78 beispielsweise für die Anzeige auf 123457 gerundet werden, für alle Berechnungen wird jedoch der ursprüngliche, ungerundete Wert verwendet.

Jahrhundertbereich für 2-stellige Jahreszahlen. Hiermit wird der Bereich der Jahre für zweistellig eingegebene und/oder angezeigte Variablen im Datumsformat definiert (zum Beispiel 10/28/86, 29-OKT-87). In der automatischen Einstellung umfaßt der Bereich die 69 Jahre vor und die 30 Jahre nach dem aktuellen Jahr (zusammen mit dem aktuellen Jahr ergibt das 100 Jahre). Bei einem benutzerdefinierten Wert wird das letzte Jahr automatisch anhand des Werts für das erste Jahr berechnet.

Zufallszahlengenerator. Zwei verschiedene Zufallszahlengeneratoren stehen zur Verfügung:

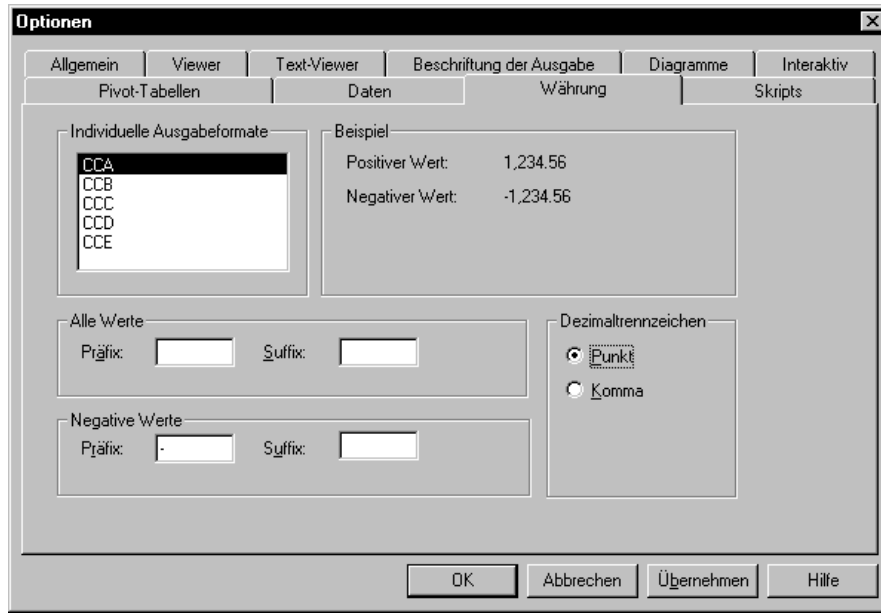
- **SPSS12-kompatibel.** Der Zufallszahlengenerator, der in SPSS 12 und älteren Versionen verwendet wird. Wenn zufallsbestimmte Ergebnisse reproduziert werden sollen, die in einer älteren Version mit einem bestimmten Startwert erzeugt wurden, ist dieser Zufallszahlengenerator zu verwenden.
- **Mersenne-Twister.** Ein neuerer Zufallszahlengenerator, der für Simulationszwecke eine höhere Zuverlässigkeit bietet. Sofern es nicht darum geht, zufallsbestimmte Ergebnisse aus SPSS 12 oder älteren Versionen zu reproduzieren, sollte man diesen Zufallszahlengenerator verwenden.

Optionen: Währung

Sie können bis zu fünf spezielle Anzeigeformate für Währungen erstellen, die über jeweils ein spezielles Präfix und Suffix verfügen und eine spezielle Behandlung negativer Werte beinhalten können.

Die fünf Namen der benutzerdefinierten Währungsfomate lauten CCA, CCB, CCC, CCD und CCE. Sie können die Namen der Formate nicht ändern und keine neuen hinzufügen. Sie können ein Währungsformat ändern, indem Sie den Formatnamen aus der Quelliste auswählen und die gewünschten Änderungen vornehmen.

Abbildung 44-9
Dialogfeld "Optionen", Registerkarte "Währung"



Die für die Währungsformate definierten Präfixe, Suffixe und Dezimalzeichen dienen nur zur Anzeige. Sie können im Daten-Editor keine Werte mit Zeichen für spezielle Währungen eingeben.

So erstellen Sie Währungsformate

- Klicken Sie auf die Registerkarte Währung.

- Wählen Sie eines der Währungsformate (CCA, CCB, CCC, CCD oder CCE) aus der Liste aus.
- Geben Sie das Präfix, das Suffix und ein Dezimaltrennzeichen ein.
- Klicken Sie auf OK oder Zuweisen.

Optionen: Skripts

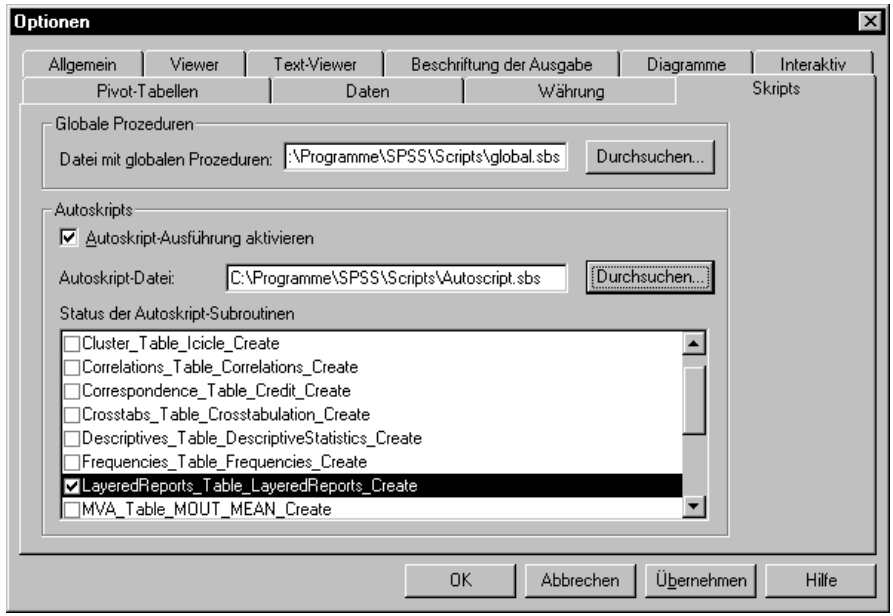
Verwenden Sie die Registerkarte “Skripts”, um die Datei für globale Prozeduren und die Autoskript-Datei anzugeben, und wählen Sie hier die gewünschten Autoskript-Subroutinen aus. Sie können Skripts zum Automatisieren vieler Funktionen verwenden, beispielsweise zum Anpassen von Pivot-Tabellen.

Globale Prozeduren. Eine Datei für globale Prozeduren ist eine Bibliothek von Skript-Subroutinen und Funktionen, die durch Skript-Dateien, einschließlich der Autoskript-Datei, aufgerufen werden können.

Anmerkung: In der Voreinstellung ist die mit dem Programm ausgelieferte Datei für globale Prozeduren ausgewählt. Viele der verfügbaren Skripts nutzen Funktionen und Subroutinen aus dieser Datei für globale Prozeduren. Diese Skripts funktionieren nicht, wenn Sie eine andere Datei für globale Prozeduren festlegen.

Autoskripts. Eine Autoskript-Datei ist eine Auswahl von Autoskript-Subroutinen, die jedesmal automatisch aufgerufen werden, wenn Sie Prozeduren ausführen, die bestimmte Arten von Ausgabeobjekten erzeugen.

Abbildung 44-10
Dialogfeld "Optionen", Registerkarte "Skripts"



Alle Subroutinen der aktuellen Autoskript-Datei werden angezeigt. Dadurch können Sie die Ausführung einzelner Subroutinen aktivieren oder deaktivieren.

So legen Sie die Optionen für Autoskripts und globale Prozeduren fest:

- ▶ Klicken Sie auf die Registerkarte Skripts.
- ▶ Wählen Sie Autoskript-Ausführung aktivieren aus.
- ▶ Wählen Sie die gewünschten Autoskript-Subroutinen aus.

  Sie können auch eine andere Autoskript-Datei oder eine andere Datei für globale Prozeduren festlegen.

Kapitel
45

Anpassen von Menüs und Symbolleisten

Menü-Editor

Sie können den Menü-Editor zum Anpassen der Menüs von SPSS verwenden. Mit dem Menü-Editor stehen Ihnen die folgenden Möglichkeiten zur Verfügung:

- Sie können Einträge zu Menüs hinzufügen, mit denen angepaßte SPSS-Skripts ausgeführt werden.
- Sie können Einträge zu Menüs hinzufügen, mit denen SPSS-Befehlssyntax-Dateien ausgeführt werden.
- Sie können Einträge zu Menüs hinzufügen, mit denen andere Anwendungen gestartet und Daten aus SPSS automatisch an andere Anwendungen übergeben werden.

Sie können Daten in den folgenden Formaten an andere Anwendungen versenden: SPSS, Excel 4.0, Lotus 1-2-3 Version 3, SYLK, durch Tabulatorzeichen getrennt, und dBASE IV.

So fügen Sie den Menüs von SPSS Einträge hinzu:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

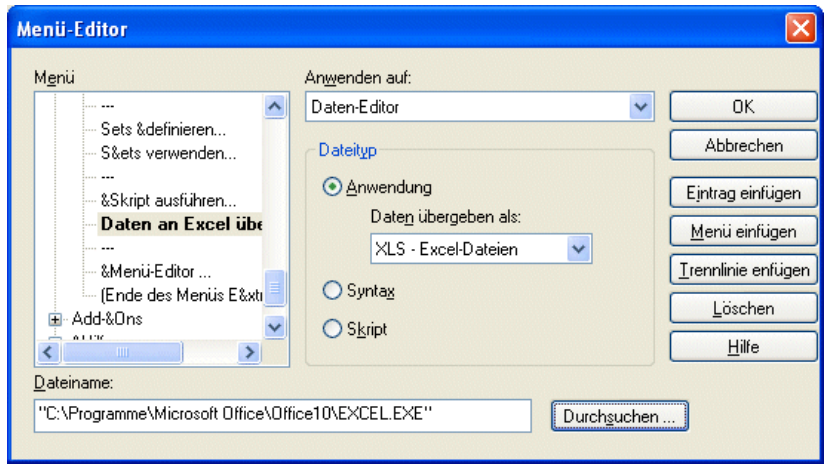
Extras
  Menü-Editor...

▶ Doppelklicken Sie im Dialogfeld “Menü-Editor” auf das Menü, dem Sie einen neuen Eintrag hinzufügen möchten.

▶ Wählen Sie den Menüeintrag aus, über dem der neue Eintrag eingefügt werden soll.

- Klicken Sie zum Einfügen des neuen Menüeintrags auf Eintrag einfügen.
- Wählen Sie den Dateityp für den neuen Eintrag aus. Hierbei haben Sie die Auswahl aus Skriptdatei, Befehlssyntax-Datei und externer Anwendung.
- Klicken Sie auf Durchsuchen, und wählen Sie die Datei aus, die dem Menüeintrag zugewiesen werden soll.

Abbildung 45-1
Dialogfeld "Menü-Editor"



Sie können außerdem vollständig neue Menüs erstellen und Trennlinien zwischen Menüeinträgen einfügen.

Wahlweise können Sie festlegen, daß der Inhalt des Daten-Editors automatisch an eine andere Anwendung übergeben wird, wenn Sie diese Anwendung aus dem Menü auswählen.

Anpassen von Symbolleisten

Sie können die Symbolleisten von SPSS anpassen und neue Symbolleisten erstellen. Symbolleisten können Symbole für alle in SPSS verfügbaren Funktionen enthalten. Dies schließt Symbole für alle über Menüs verfügbaren Aktionen ein. Außerdem können Symbole enthalten sein, mit denen andere Anwendungen gestartet sowie Befehlssyntax-Dateien und Skriptdateien ausgeführt werden.

Symbolleisten anzeigen

Im Dialogfeld “Symbolleisten anzeigen” können Symbolleisten ein- bzw. ausgeblendet und angepaßt sowie neue Symbolleisten erstellt werden. Symbolleisten können Symbole für alle in SPSS verfügbaren Funktionen enthalten. Dies schließt Symbole für alle über Menüs verfügbaren Aktionen ein. Außerdem können Symbole enthalten sein, mit denen andere Anwendungen gestartet sowie Befehlssyntax-Dateien und Skriptdateien ausgeführt werden.

Abbildung 45-2
Dialogfeld “Symbolleisten anzeigen”

So passen Sie Symbolleisten an:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Ansicht
  Symbolleisten...

- Wählen Sie die Symbolleiste aus, die Sie anpassen möchten, und klicken Sie auf Anpassen. Klicken Sie zum Erstellen einer neuen Symbolleiste auf Neue Symbolleiste.

- Geben Sie bei einer neuen Symbolleiste einen Namen für die Symbolleiste ein, wählen Sie die Fenster aus, in denen die Symbolleiste angezeigt werden soll, und klicken Sie auf Anpassen.

- Wenn Sie aus der Liste “Kategorien” eine Kategorie auswählen, werden die in dieser Kategorie verfügbaren Symbole angezeigt.
- Ziehen Sie die gewünschten Symbole auf die im Dialogfeld angezeigte Symbolleiste.
- Zum Entfernen eines Symbols aus einer Symbolleiste ziehen Sie das Symbol aus der im Dialogfeld angezeigten Symbolleiste.

So erstellen Sie ein Symbol zum Öffnen einer Datei oder zum Ausführen einer Befehlssyntax-Datei bzw. eines Skripts:

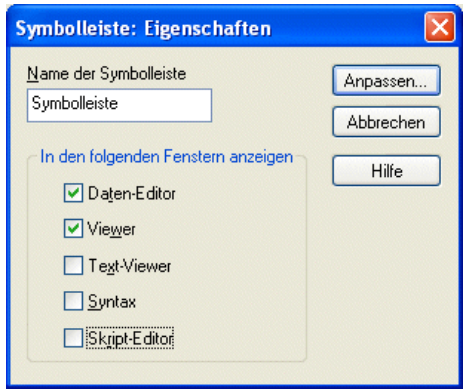
- Klicken Sie im Dialogfeld “Symbolleiste anpassen” auf Neues Symbol.
- Geben Sie eine aussagekräftige Beschriftung für das Symbol ein.
- Wählen Sie die gewünschte Aktion für das Symbol aus, also das Öffnen einer Datei oder das Ausführen einer Befehlssyntax-Datei bzw. eines Skripts.
- Klicken Sie auf Durchsuchen, und wählen Sie die Datei oder die Anwendung aus, die dem Symbol zugeordnet werden soll.

Neue Symbole werden in der Kategorie “Benutzerdefiniert” angezeigt. Hier finden Sie außerdem die benutzerdefinierten Menüeinträge.

Symbolleiste: Eigenschaften

Verwenden Sie das Dialogfeld “Symbolleiste: Eigenschaften”, um auszuwählen, in welchen Fenstern die ausgewählte Symbolleiste angezeigt werden soll. In diesem Dialogfeld können außerdem Namen für neue Symbolleisten eingegeben werden.

Abbildung 45-3
Dialogfeld "Symbolleiste: Eigenschaften"



So legen Sie die Eigenschaften von Symbolleisten fest:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Ansicht
  Symbolleisten...

- Klicken Sie bei vorhandenen Symbolleisten erst auf Anpassen und dann im Dialogfeld "Symbolleiste anpassen" auf Eigenschaften.

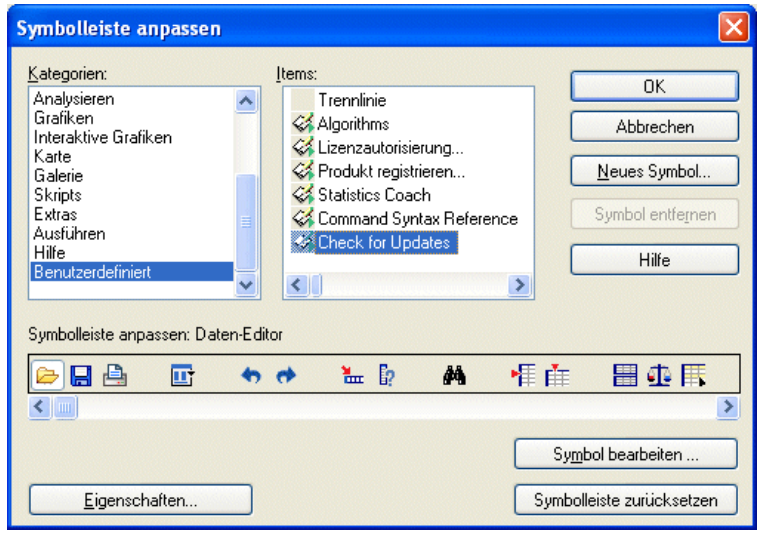
- Klicken Sie bei einer neuen Symbolleiste auf Neues Symbol.

- Wählen Sie die Fenster aus, in denen die Symbolleiste angezeigt werden soll. Geben Sie für eine neue Symbolleiste außerdem einen Namen ein.

Symbolleiste anpassen

Verwenden Sie das Dialogfeld "Symbolleiste anpassen" zum Anpassen von vorhandenen Symbolleisten und zum Erstellen von neuen Symbolleisten. Symbolleisten können Symbole für alle in SPSS verfügbaren Funktionen enthalten. Dies schließt Symbole für alle über Menüs verfügbaren Aktionen ein. Außerdem können Symbole enthalten sein, mit denen andere Anwendungen gestartet sowie Befehlssyntax-Dateien und Skriptdateien ausgeführt werden.

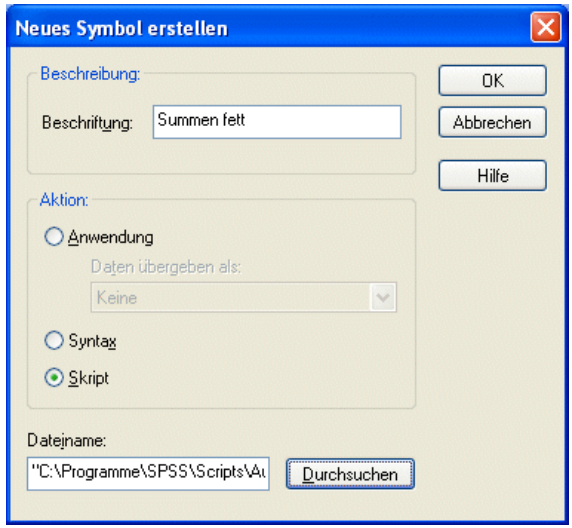
Abbildung 45-4
Dialogfeld "Symbolleiste anpassen"



Neues Symbol erstellen

Verwenden Sie das Dialogfeld "Neues Symbol erstellen" zum Erstellen von Symbolen, mit denen Sie andere Anwendungen starten und Befehlssyntax-Dateien sowie Skriptdateien ausführen können.

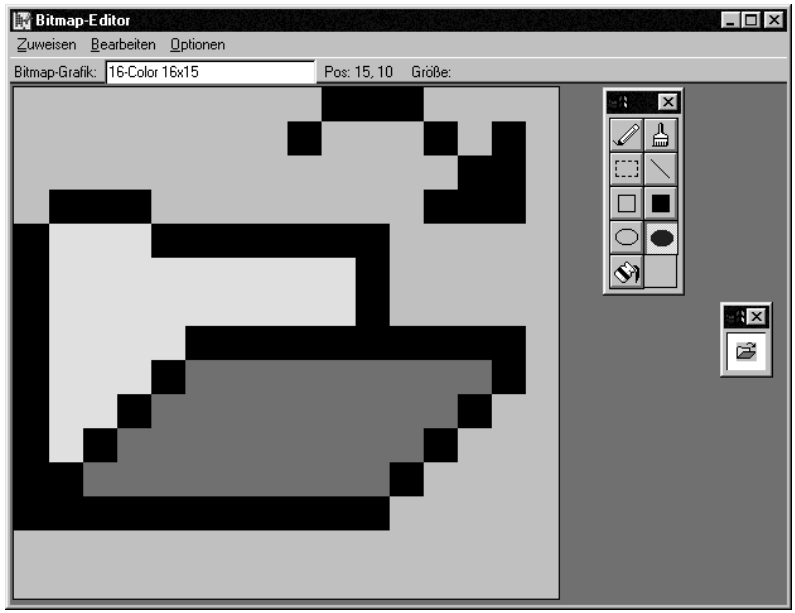
Abbildung 45-5
Dialogfeld "Neues Symbol erstellen"



Bitmap-Editor für Symbolleisten

Verwenden Sie den Bitmap-Editor zum Erstellen von eigenen Symbolen für die Schaltflächen von Symbolleisten. Dies ist insbesondere für die Symbole nützlich, mit denen Sie Skripts und Syntax ausführen sowie andere Anwendungen starten.

Abbildung 45-6
Bitmap-Editor



So bearbeiten Sie Bitmaps für Symbolleisten

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Ansicht
  Symbolleisten...

- Wählen Sie die Symbolleiste aus, die Sie bearbeiten möchten, und klicken Sie auf Anpassen.

- Klicken Sie in der Vorschau der Symbolleiste auf das Symbol mit der Bitmap-Grafik, die Sie bearbeiten möchten.

- Klicken Sie auf Symbol bearbeiten.

- Verwenden Sie die Symbolleiste und die Farbpalette zum Bearbeiten der Bitmap-Grafik oder zum Erstellen einer neuen Bitmap-Grafik.

Kapitel
46

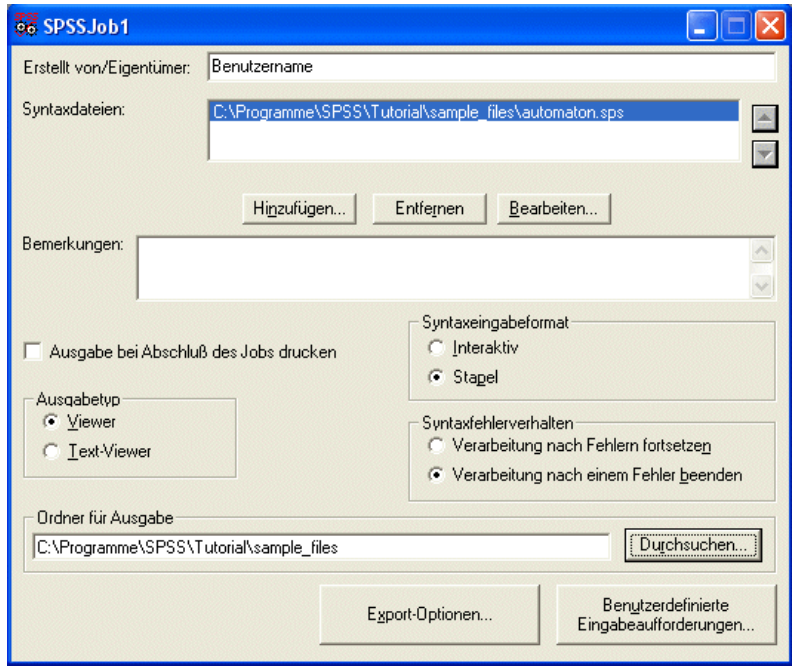
SPSS-Produktionsmodus

Mit dem SPSS-Produktionsmodus kann das Programm Berechnungen automatisch ausführen. Das Programm läuft dann bedienungsfrei ab und wird nach Ausführen des letzten Befehls beendet, so daß Sie sich indessen anderen Aufgaben widmen können. Der Produktionsmodus bietet sich an, wenn Sie oft dieselben zeitaufwendigen Analysen durchführen müssen, beispielsweise für wöchentliche Berichte.

Im SPSS-Produktionsmodus werden die Befehle über Befehlssyntax-Dateien an das Programm übermittelt. Dabei handelt es sich um einfache Textdateien für Befehlssyntax. Zum Erstellen der Datei können Sie einen beliebigen Text-Editor verwenden. Sie können Befehlssyntax auch erstellen, indem Sie die in einem Dialogfeld getroffene Auswahl in ein Syntax-Fenster einfügen oder die Journaldatei bearbeiten.

Nachdem Sie Syntaxdateien erstellt und in einen Produktionsjob aufgenommen haben, können Sie diese im Produktionsmodus anzeigen lassen und bearbeiten.

Abbildung 46-1
SPSS-Produktionsmodus



Syntaxeingabeformat. Steuert die Form der Syntaxregeln für den Job:

- **Interaktiv.** Jeder Befehl muß mit einem Punkt enden. Punkte können an einer beliebigen Stelle im Befehl auftreten, und Befehle können sich über mehrere Zeilen erstrecken. Ein Punkt als letztes Zeichen in einer Zeile wird jedoch stets als Ende des Befehls interpretiert. Fortgesetzte Zeilen und neue Befehle können an einer beliebigen Position in einer neuen Zeile beginnen. Diese “interaktiven” Regeln sind in Kraft, wenn Sie Befehle in einem Syntaxfenster auswählen und ausführen.
- **Stapel.** Jeder Befehl muß am Anfang einer neuen Zeile beginnen (Leerzeichen vor dem Befehl sind nicht zulässig); fortgesetzte Zeilen müssen um mindestens ein Leerzeichen eingerückt sein. Sollen neue Befehle eingerückt werden, geben Sie ein Pluszeichen, einen Bindestrich oder einen Punkt als erstes Zeichen am Anfang der Zeile ein, und rücken Sie dann den eigentlichen Befehl nach Wunsch ein. Der Punkt am Ende des Befehls ist optional. Diese Einstellung ist kompatibel mit den Syntaxregeln für Befehlsdateien im Befehl `INCLUDE`.

Syntaxfehlerverhalten. Steuert die Behandlung von Fehlerbedingungen im Job:

- **Weiter.** Fehler im Job führen nicht automatisch dazu, daß die Befehlsverarbeitung abgebrochen wird. Die Befehle in den Produktionsjob-Dateien werden als Teil des normalen Befehlsstroms behandelt, und die Befehlsverarbeitung wird normal fortgesetzt.
- **Stopp.** Die Befehlsverarbeitung wird angehalten, sobald der erste Fehler in einer Produktionsjob-Datei auftritt. Diese Einstellung ist kompatibel mit dem Verhalten von Befehlsdateien im Befehl `INCLUDE`.

Ergebnisse von Produktionsjobs. Bei jedem Produktionslauf wird eine Ausgabedatei mit dem Namen des Produktionsjobs und der Dateinamenerweiterung *.spo* angelegt. So wird zum Beispiel für eine Produktionsjob-Datei mit dem Namen *prodjob.spp* eine Ausgabedatei mit dem Namen *prodjob.spo* angelegt. Die Ausgabedatei ist ein Viewer-Dokument.

Ausgabetyp. Bei Ausgabe im Viewer werden Pivot-Tabellen und hochauflösende, interaktive Diagramme erzeugt. Bei Ausgabe im Text-Viewer werden Textausgaben und Metadatei-Grafiken von Diagrammen erzeugt. Im Text-Viewer können zwar Textausgaben, jedoch keine Diagramme bearbeitet werden.

Verwenden des Produktionsmodus

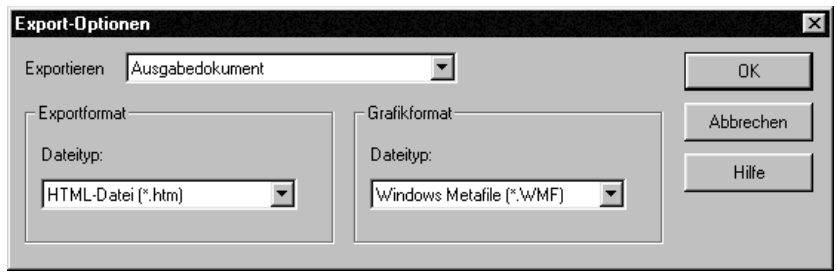
- ▶ Legen Sie eine Befehlssyntax-Datei an.
- ▶ Starten Sie den SPSS-Produktionsmodus über das Startmenü von Windows.
- ▶ Geben Sie die Syntaxdateien an, die Sie im Produktionsjob verwenden möchten. Klicken Sie zum Auswählen der Syntaxdateien auf Hinzufügen.
- ▶ Speichern Sie die Produktionsjob-Datei.
- ▶ Starten Sie die Produktionsjob-Datei. Klicken Sie auf der Symbolleiste auf die Schaltfläche Ausführen, oder wählen Sie die folgenden Befehle aus den Menüs aus:

  Ausführen
  Produktionsjob

Export-Optionen

Im Dialogfeld “Export-Optionen” können Sie festlegen, daß Pivot-Tabellen und Textausgaben im HTML-, Text-, Word/RTF- oder Excel-Format gespeichert werden. Sie können auch festlegen, daß Diagramme in verschiedenen gebräuchlichen Formaten gespeichert werden, die auch von anderen Anwendungen verwendet werden.

Abbildung 46-2
Dialogfeld “Export-Optionen”



Export

In dieser Dropdown-Liste wird angegeben, was exportiert werden soll.

Ausgabedokument. Hiermit können Sie beliebige Kombinationen von Pivot-Tabellen, Textausgaben und Diagrammen exportieren.

- Beim HTML- und Textformat werden Diagramme im derzeit ausgewählten Exportformat für Diagramme exportiert. Bei Dokumenten im HTML-Format werden Diagramme als Verweis eingebettet. Exportieren Sie Diagramme in diesem Fall in einem für die Aufnahme in HTML-Dokumente geeigneten Format. Bei Dokumenten im Textformat wird in der Textdatei für jedes Diagramm eine Zeile mit der Angabe des Dateinamens für das exportierte Diagramm eingefügt.
- Beim Word/RTF-Format werden Diagramme im Windows-Metafile-Format exportiert und in das Word-Dokument eingebettet.
- In Excel-Dokumente können keine Diagramme exportiert werden.

Ausgabedokument (ohne Diagramme). Hiermit werden Pivot-Tabellen und Textausgaben exportiert. Sämtliche Diagramme im Viewer werden ignoriert.

Nur Diagramme. Hiermit werden nur Diagramme exportiert. Bei Dokumenten im HTML- und Textformat sind folgende Exportformate gültig: Enhanced Metafile (EMF) Windows-Metadateien (WMF), Windows-Bitmaps (BMP), Encapsulated PostScript (EPS), JPEG, TIFF, PNG und Macintosh-PICT. Bei Word/RTF-Dokumenten werden Diagramme stets im Windows-Metafile-Format exportiert.

Exportformat

Bei Ausgabedokumenten sind die folgenden Optionen verfügbar: HTML, Text, Word/RTF und Excel. Beim HTML- und Textformat werden Diagramme im derzeit ausgewählten Diagrammformat exportiert. Wenn Sie die Option "Nur Diagramme" ausgewählt haben, wählen Sie ein Exportformat aus der Dropdown-Liste aus. Bei Ausgabedokumenten werden Pivot-Tabellen und Text folgendermaßen exportiert:

- **HTML-Datei (*.htm).** Pivot-Tabellen werden als HTML-Tabellen exportiert. Textausgaben werden als vorformatierter HTML-Text exportiert.
- **Textdatei (*.txt).** Pivot-Tabellen können als durch Tabulatoren getrennter Text oder als durch Leerzeichen getrennter Text exportiert werden. Alle Textausgaben werden in durch Leerzeichen getrenntem Format exportiert.
- **Excel-Datei (*.xls).** Die Zeilen, Spalten und Zellen von Pivot-Tabellen werden mit sämtlichen Formatierungsattributen wie Zellenrahmen, Schriftarten, Hintergrundfarben usw. als Excel-Zeilen, -Spalten und -Zellen exportiert. Textausgaben werden mit allen Schriftartattributen exportiert. Jede Zeile in der Textausgabe entspricht einer Zeile in der Excel-Datei, wobei der gesamte Inhalt der Zeile in einer einzelnen Zelle enthalten ist.
- **Word/RTF-Datei (*.doc).** Pivot-Tabellen werden mit sämtlichen Formatierungsattributen wie Zellenrahmen, Schriftarten, Hintergrundfarben usw. als Word-Tabellen exportiert. Textausgaben werden als formatierter RTF-Text exportiert. Textausgaben in SPSS werden immer mit einem nicht proportionalen Zeichensatz (mit festem Abstand) angezeigt und mit denselben Schriftartenattributen exportiert. Für die richtige Ausrichtung von durch Leerzeichen getrennten Textausgaben ist ein nicht proportionaler Zeichensatz (mit festem Abstand) erforderlich.

Bildformat

Über das Bildformat legen Sie das Exportformat für Diagramme fest. Sie können Diagramme in den folgenden Formaten speichern: Enhanced Metafile (EMF) Windows-Metadateien (WMF), Windows-Bitmaps (BMP), Encapsulated PostScript (EPS), JPEG, TIFF, PNG oder Macintosh-PICT.

Die Namen für exportierte Diagramme basieren auf dem Dateinamen des Produktionsjobs, einer laufenden Nummer und der Dateinamenerweiterung des ausgewählten Formats. Wenn beispielsweise der Produktionsjob *prodjob.spp* Diagramme im Windows-Metafile-Format exportiert, lauten die Namen der Diagramme *prodjob1.wmf, prodjob2.wmf, prodjob3.wmf* usw.

Optionen für Text und Bilder

Die Optionen für das Exportieren von Text (beispielsweise durch Tabulatoren oder Leerzeichen getrennt) und Diagrammen (beispielsweise Farbeinstellungen, Größe und Auflösung) werden in SPSS festgelegt und können im Produktionsmodus nicht geändert werden. Verwenden Sie zum Ändern der Optionen für den Export von Texten und Diagrammen den Befehl “Exportieren” im Menü “Datei” von SPSS.

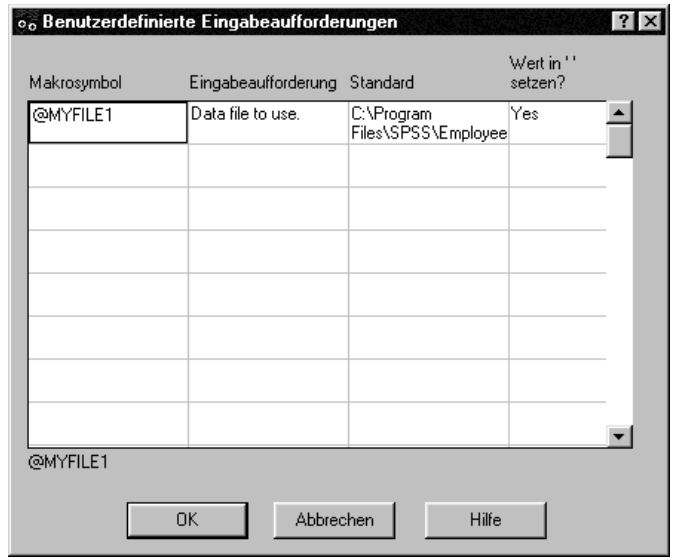
Exportieren aus dem Text-Viewer

Bei Text-Viewer-Ausgaben kann die Ausgabe nur in einfachem Textformat exportiert werden. Diagramme für die Text-Viewer-Ausgabe können nicht exportiert werden.

Benutzerdefinierte Eingabeaufforderungen

In Produktionsjob-Dateien definierte und in Befehlssyntax-Dateien verwendete Makrosymbole vereinfachen Aufgaben, wie zum Beispiel das Durchführen derselben Analyse für mehrere Datendateien oder das Ausführen derselben Befehle für mehrere Variablen-Sets. So können Sie beispielsweise das Makrosymbol *@datdtei* definieren, sodass bei jedem Ausführen von Produktionsjobs, die anstelle eines Namens für die Datendatei in der Befehlssyntax-Datei den String *@datdtei* enthalten, ein Name für die Datendatei abgefragt wird.

Abbildung 46-3
Dialogfeld "Benutzerdefinierte Eingabeaufforderungen"



Makrosymbol. Der in der Befehlssyntax-Datei verwendete Makroname zum Aufrufen des Makros, mit dem der Benutzer zum Eingeben von Informationen aufgefordert wird. Der Name des Makrosymbols muß mit einem @ beginnen. *Anmerkung*: Diese "Makro"-Symbole stehen in keinerlei Beziehung zu den Makros, die mit `DEFINE-!ENDDEFINE` in den SPSS-Makrofunktionen erstellt werden.

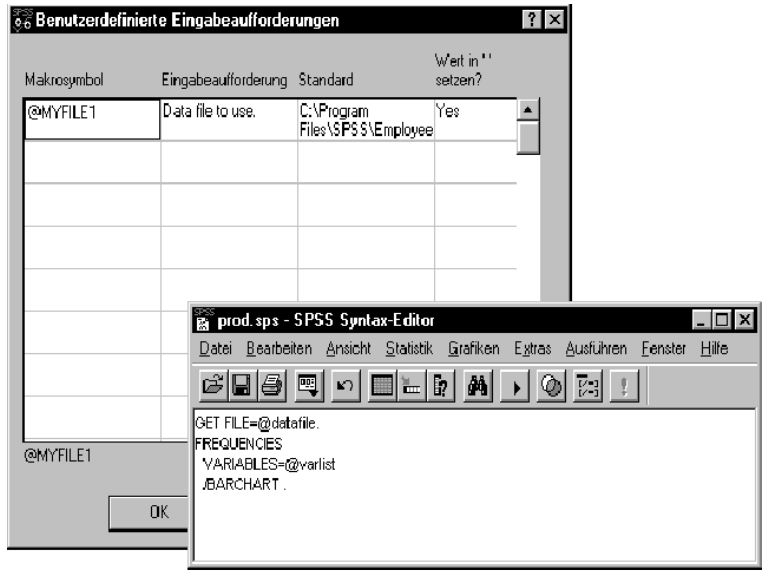
Eingabeaufforderung. Der beschreibende Text, der angezeigt werden soll, wenn der Produktionsjob zur Eingabe von Informationen auffordert. Sie können zum Beispiel mit der Formulierung "Welche Datendatei möchten Sie verwenden?" ein Feld bezeichnen, in das der Name einer Datendatei eingegeben werden muß.

Standard. Dies ist der Wert, der vom Produktionsjob in der Standardeinstellung vorgegeben wird, falls Sie keinen anderen Wert eingeben. Dieser Wert wird angezeigt, wenn Sie beim Ausführen des Produktionsjobs zum Eingeben von Informationen aufgefordert werden. Sie können den Wert während der Ausführung des Programms ersetzen oder modifizieren.

Wert in Anführungszeichen setzen? Geben Sie J oder Ja ein, wenn der Wert in Anführungszeichen eingeschlossen werden soll. Lassen Sie das Feld andernfalls leer, oder geben Sie N oder Nein ein. Bei der Angabe von Dateinamen müssen Sie

beispielsweise Ja eingeben, weil Dateinamen in Anführungszeichen eingeschlossen werden müssen.

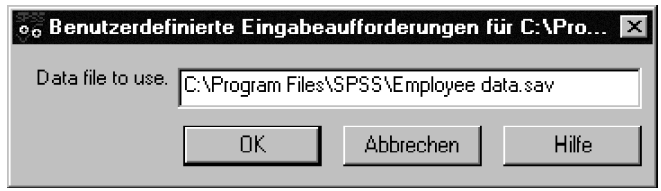
Abbildung 46-4
Makro-Eingabeaufforderungen in einer Befehlssyntax-Datei



Makro-Eingabeaufforderungen im Produktionsmodus

Der SPSS-Produktionsmodus fordert Sie zur Eingabe von Werten auf, wenn Sie einen Produktionsjob mit definierten Makrosymbolen ausführen. Sie können die angezeigten Standardwerte ersetzen oder ändern. Diese Werte ersetzen dann die Makrosymbole in allen zu dem Produktionsjob gehörenden Befehlssyntax-Dateien.

Abbildung 46-5
Dialogfeld einer Makro-Eingabeaufforderung im Produktionsmodus



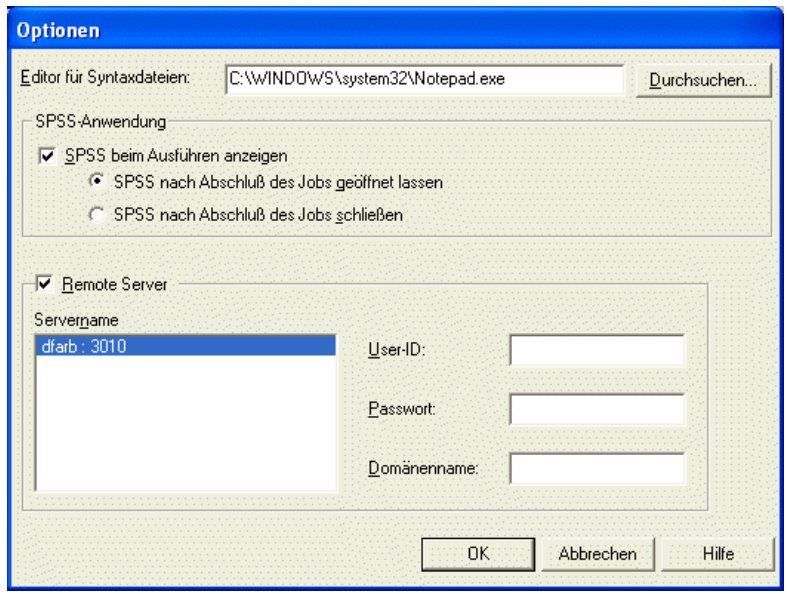
Anmerkung: Diese “Makro”-Symbole stehen in keinerlei Beziehung zu den Makros, die mit `DEFINE-!ENDDEFINE` in den SPSS-Makrofunktionen erstellt werden. Der Produktionsjob fordert Sie zur Eingabe dieser Werte auf, auch wenn keine der Befehlssyntax-Dateien im Job Verweise auf die definierten “Makros” enthält.

Produktionsmodus: Optionen

Für den Produktionsmodus können Sie folgende Optionen festlegen:

- Sie können einen Standard-Text-Editor für Syntaxdateien festlegen, der geöffnet wird, wenn Sie im Hauptdialogfeld auf die Schaltfläche “Bearbeiten” klicken.
- Sie können festlegen, daß der Produktionsjob als verborgener Prozeß im Hintergrund ausgeführt wird oder die berechneten Ergebnisse während der Ausführung des Jobs angezeigt werden.
- Sie können einen Remote-Server, einen Domänennamen, einen Benutzernamen und ein Paßwort für verteilte Analysen angeben (nur verfügbar, wenn Sie über ein Netzwerk auf die Server-Version von SPSS zugreifen können). Wenn Sie diese Einstellungen nicht vornehmen, werden die Standardeinstellungen aus dem Dialogfeld “Login beim SPSS-Server” verwendet. Sie können nur Remote-Server auswählen, die bereits im Dialogfeld “SPSS-Server hinzufügen” (Menü “Datei”, Befehl “Server umschalten”, Schaltfläche “Hinzufügen”) definiert wurden.

Abbildung 46-6
Dialogfeld "Optionen"



Ändern der Optionen für den Produktionsmodus

Wählen Sie folgende Befehle aus den Menüs des Produktionsmodus aus:

Bearbeiten
  Optionen...

Festlegen von Formaten für Produktionsjobs

Eine Reihe von Einstellungen in SPSS gewährleisten, daß für die in Produktionsjobs angelegten Pivot-Tabellen das am besten geeignete Format verwendet wird:

Tabellenvorlagen. Durch Bearbeiten und Speichern von Tabellenvorlagen (im Menü "Format" einer aktivierten Pivot-Tabelle) können Sie viele Eigenschaften von Pivot-Tabellen festlegen. Sie können Schriftgrößen und -arten, Farben und Rahmen festlegen. Sie können die Aufteilung breiter Tabellen auf mehrere Seiten verhindern, indem Sie im Dialogfeld "Tabelleneigenschaften" auf der Registerkarte "Drucken" die Option Breite Tabelle auf Seitengröße verkleinern wählen.

Beschriftung der Ausgabe. Mit diesen Optionen (Menü "Bearbeiten", Befehl "Optionen", Registerkarte "Beschriftung der Ausgabe") können Sie die Anzeige von Informationen zu Variablen und Datenwerten in Pivot-Tabellen festlegen. Sie können Variablennamen und/oder definierte Variablenlabels sowie tatsächliche Datenwerte und/oder definierte Wertlabels anzeigen lassen. Aussagekräftige Variablen- und Wertelabels vereinfachen meist die Interpretation der Ergebnisse. In manchen Tabellen können lange Labels jedoch eher stören.

Spaltenbreite. Mit den Optionen für Pivot-Tabellen (Menü "Bearbeiten", Befehl "Optionen", Registerkarte "Pivot-Tabelle") legen Sie Einstellungen für die Standard-Tabellenvorlage und die automatische Anpassung der Spaltenbreiten in Pivot-Tabellen fest.

- **Beschriftungen.** Hiermit passen Sie die Spaltenbreite an die Breite der Spaltenbeschriftung an. Dadurch erhalten Sie kompaktere Tabellen, aber Datenwerte, die breiter als die Beschriftung sind, werden nicht angezeigt. (Sternchen weisen auf Werte hin, die zu breit für die Anzeige sind.)
- **Beschriftungen und Daten.** Hiermit passen Sie die Spaltenbreite an die Spaltenbeschriftung oder den Datenwert an, je nachdem, welcher der beiden länger ist. Dies erzeugt breitere Tabellen, stellt jedoch die Anzeige aller Werte sicher.

In Produktionsjobs werden die aktuelle Tabellenvorlage und die aktuell gültigen Optionseinstellungen verwendet. Sie können die Tabellenvorlage und die gewünschten Optionen vor dem Ausführen des Produktionsjobs einstellen. Sie können dafür jedoch auch den Befehl `SET` in Ihren Syntaxdateien verwenden. Bei der Verwendung des Befehls `SET` in Syntaxdateien können Sie für einen Job gleichzeitig mehrere Einstellungen für Tabellenvorlagen und Optionen vornehmen.

Anlegen einer benutzerdefinierten Standard-Tabellenvorlage

- ▶ Aktivieren Sie die Pivot-Tabelle, indem Sie auf eine beliebige Stelle der Tabelle doppelklicken.
- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

  Format
  Tabellenvorlagen...

- ▶ Wählen Sie eine Tabellenvorlage aus der Liste aus, und klicken Sie auf Tabellenvorlage bearbeiten.

- Passen Sie die Tabelleneigenschaften an die gewünschten Attribute an.
- Speichern Sie die Tabellenvorlage, indem Sie erst auf Vorlage speichern oder Speichern unter und dann auf OK klicken.
- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Bearbeiten
  Optionen...
- Klicken Sie auf die Registerkarte Pivot-Tabellen.
- Wählen Sie die Tabellenvorlage aus der Liste aus, und klicken Sie dann auf OK.

Festlegen von Optionen für Produktionsjobs

- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Bearbeiten
  Optionen...
- Wählen Sie die gewünschten Optionen aus.
- Klicken Sie auf OK.

  Im Dialogfeld "Optionen" können Sie die Standard-Tabellenvorlage festlegen sowie Einstellungen für Ausgabebeschriftungen und die automatische Anpassung der Spaltenbreite vornehmen. Die Einstellungen der Optionen werden vom Programm gespeichert. Beim Ausführen eines Produktionsjobs werden die beim letzten Aufrufen des Programms wirksamen Einstellungen verwendet.

Festlegen des Formats von Pivot-Tabellen mit Befehlssyntax

SET TLOOK. Hiermit legen Sie die Standard-Tabellenvorlage für neue Pivot-Tabellen fest, zum Beispiel:

- `SET TLOOK = 'c:\prodjobs\tabelle.tlo'.`

SET TVARS. Hiermit wird die Anzeige von Variablennamen und -labels in neuen Pivot-Tabellen festgelegt.

- `SET TVARS = LABELS`: Es werden Variablenlabels angezeigt.
- `SET TVARS = NAMES`: Es werden Variablennamen angezeigt.
- `SET TVARS = BOTH`: Es werden Variablennamen und -labels angezeigt.

SET ONUMBER. Hiermit wird die Anzeige von Datenwerten und Wertelabels in neuen Pivot-Tabellen festgelegt.

- `SET ONUMBER = LABELS`: Es werden Wertelabels angezeigt.
- `SET ONUMBER = VALUES`: Es werden Datenwerte angezeigt.
- `SET ONUMBER = BOTH`: Es werden Datenwerte und Wertelabels angezeigt.

SET TFIT. Hiermit können Sie die automatische Anpassung der Spaltenbreiten für neue Pivot-Tabellen festlegen.

- `SET TFIT = LABELS`: Die Spaltenbreiten werden an die Breiten der Spaltenbeschriftungen angepaßt.
- `SET TFIT = BOTH`: Die Spaltenbreiten werden an die Breiten der Spaltenbeschriftungen oder der größten Datenwerte angepaßt, je nachdem, welches Element breiter ist.

Ausführen von Produktionsjobs aus der Befehlszeile

Mit Befehlszeilenschaltern können Sie Produktionsjobs unter Verwendung von Dienstprogrammen, wie beispielsweise dem Systemdienst aus Microsoft Plus!, zeitgesteuert ausführen. Sie können Produktionsjobs aus einer Befehlszeile heraus mit den folgenden Schaltern ausführen:

-r. Startet den Produktionsjob. Wenn der Produktionsjob benutzerdefinierte Eingabeaufforderungen enthält, müssen Sie die nötigen Informationen vor Ausführung des Produktionsjob eingeben.

-s. Führt den Produktionsjob aus und unterdrückt alle benutzerdefinierten Eingabeaufforderungen und Warnmeldungen. Dabei werden automatisch die Standardwerte für benutzerdefinierte Eingaben verwendet.

Verteilte Analyse. Wenn Sie über Netzwerkzugriff auf die Server-Version von SPSS verfügen, können Sie außerdem die folgenden Schalter verwenden, um den Produktionsmodus im Modus für verteilte Analysen auszuführen.

-x. Name oder IP-Adresse des Remote-Servers.

-n. Portnummer.

-d. Domänenname.

-u. Benutzername für den Zugriff auf den Remote-Server.

-p. Paßwort für den Zugriff auf den Remote-Server.

Wenn Sie einen der Befehlszeilenschalter für die verteilte Analyse angeben, müssen Sie auch alle weiteren Schalter für die verteilte Analyse (`-x`, `-n`, `-d`, `-u` und `-p`) angeben.

Geben Sie den vollständigen Pfadnamen für den Produktionsmodus (*spssprod.exe*) und den Produktionsjob an. Beide müssen in Anführungszeichen eingeschlossen werden, wie das folgende Beispiel zeigt:

```
"c:\programme\spss\spssprod.exe" "c:\spss\jobs\prodjob.spp" -s -r
```

Auf Schalter, für die eine weitere Angabe erforderlich ist, müssen unmittelbar ein Gleichheitszeichen und die betreffende Angabe folgen. Wenn die Angabe Leerzeichen enthält (beispielsweise der Name eines Servers in zwei Wörtern), muß die Angabe in Anführungszeichen oder Hochkommata eingeschlossen werden:

```
-x="HAL 9000" -u="geheimes Wort"
```

Standardserver. Wenn Sie über Netzwerkzugriff auf die Server-Version von SPSS verfügen, entsprechen der Standard-Server und relevante Informationen (sofern nicht mit den Befehlszeilenschaltern angegeben) dem Standard-Server, der im Dialogfeld "Login beim SPSS-Server" angegeben ist. Wenn dort kein Standard-Server angegeben ist, wird der Job im Modus für lokale Analysen ausgeführt.

Wenn Sie einen Produktionsjob im Modus für lokale Analysen ausführen möchten, Ihr Computer aber nicht der Standard-Server ist, geben Sie für alle Schalter für die verteilte Analyse leere Strings an:

```
"c:\programme\spss\spssprod.exe" "c:\spss\jobs\prodjob.spp" -x="" -n="" -d="" -u="" -p=""
```

Ausführen mehrerer Produktionsjobs

Wenn Sie mehrere Produktionsjobs mit Hilfe einer Stapelverarbeitungsdatei (.BAT) oder einem ähnlichen Verfahren ausführen, steuern Sie die Ausführung der einzelnen Jobs mit dem Befehl `Start` und dem Schalter `/wait`. So wird vermieden, daß nachfolgende Jobs schon gestartet werden, bevor der vorangegangene Jobs beendet ist:

```
cd \programme\spss
start /wait spssprod.exe prodjob1.spp -s
start /wait spssprod.exe prodjob2.spp -s
```

Im Web veröffentlichen

Mit der Option "Im Web veröffentlichen" exportieren Sie Ausgaben für die Veröffentlichung auf einem SmartViewer-Web-Server. Tabellen und Berichte, die mit SmartViewer veröffentlicht werden, können über das Internet in Echtzeit mit einem herkömmlichen Browser angezeigt und bearbeitet werden.

- Pivot-Tabellen werden als dynamische Tabellen veröffentlicht. Sie können über das Internet geändert werden und ermöglichen so verschiedene Darstellungen der Daten.
- Diagramme werden als JPEG- oder PNG-Grafikdateien veröffentlicht.
- Textausgaben werden als vorformatierter HTML-Text veröffentlicht. (Die meisten Web-Browser verwenden für vorformatierten Text in der Standardeinstellung einen Zeichensatz mit festem Abstand.)

Veröffentlichen. Hiermit können Sie die Ausgabe angeben, die Sie veröffentlichen möchten:

- **Ausgabedokument.** Das gesamte Ausgabedokument einschließlich der ausgeblendeten oder reduzierten Objekte wird veröffentlicht.
- **Ausgabedokument (ohne Anmerkungen).** Veröffentlicht alles außer den für jede Prozedur automatisch erzeugten Anmerkungstabellen.
- **Nur Tabellen.** Diagramme werden ausgeschlossen. Alle Pivot-Tabellen und alle Text-Tabellen werden veröffentlicht.
- **Nur Tabellen (ohne Anmerkungen).** Diagramme und Anmerkungstabellen werden ausgeschlossen.
- **Nur Diagramme.** Es werden nur die Diagramme im Dokument veröffentlicht.
- **Nichts.** Das Veröffentlichen im Internet wird deaktiviert. Da alle Einstellungen zusammen mit dem Produktionsjob (*SPP*-Datei) gespeichert werden, werden die Ergebnisse bei jeder Ausführung des Produktionsjobs veröffentlicht, es sei denn, Sie haben die Option Nichts ausgewählt. Die Veröffentlichung wird zwar deaktiviert, es werden jedoch weiterhin die anderen Ausgabetypen erzeugt (Viewer-Dateien, HTML-Dateien), die für den Produktionsjob festgelegt wurden.

Tabellen veröffentlichen als. Steuert, wie Pivot-Tabellen veröffentlicht werden:

- **Interaktiv.** Die Tabellen liegen als dynamische Objekte vor, die über das Internet geändert werden können und so verschiedene Darstellungen der Daten zulassen.
- **Statisch.** Die Tabellen sind statisch und können nach dem Veröffentlichen nicht mehr geändert werden.

Konfigurieren. Öffnet die Seite zum Konfigurieren der automatischen Veröffentlichung des SmartViewer-Web-Servers in einem Browserfenster. Dies ist zum Erstellen eines neuen Produktionsjobs erforderlich, der im Internet veröffentlicht werden soll.

Für den Zugriff auf den SmartViewer-Web-Server sind ein Benutzername und ein Passwort erforderlich. Wenn Sie einen neuen Produktionsjob zur Veröffentlichung im Internet erstellen, werden Sie aufgefordert, Ihren Benutzernamen und Ihr Passwort anzugeben. Diese Informationen werden in verschlüsselter Form in dem Produktionsjob gespeichert.

Anmerkung: Die Option “Im Web veröffentlichen” ist nur für Sites verfügbar, in denen SmartViewer-Web-Server installiert ist. Weiterhin ist zum Aktivieren dieser Funktionen ein Plug-in erforderlich. Wenden Sie sich an Ihren Systemadministrator oder Webmaster, falls Sie Anweisungen zum Herunterladen des Plug-ins benötigen. Speichern Sie die Ausgabe mit “Ausgabe exportieren” im HTML-Format, falls SmartViewer auf Ihrer Site nicht zur Verfügung steht.

SmartViewer-Web-Server: Anmeldung

Für Veröffentlichungen auf SmartViewer-Web-Server benötigen Sie einen gültigen Benutzernamen (Benutzer-ID) und ein Paßwort.

Wenden Sie sich an Ihren Systemadministrator oder Webmaster, wenn Sie weitere Informationen benötigen.

Kapitel
47

Skriptfunktionen in SPSS

Mit der Skriptfunktion können Sie Abläufe in SPSS automatisieren, darunter:

- Öffnen und Speichern von Datendateien,
- Anzeigen und Ändern von Dialogfeldern,
- Ausführen von Datentransformationen und statistischen Prozeduren unter Verwendung der Befehlssyntax,
- Exportieren von Diagrammen in verschiedenen Grafikformaten,
- Anpassen der Ausgaben im Viewer.

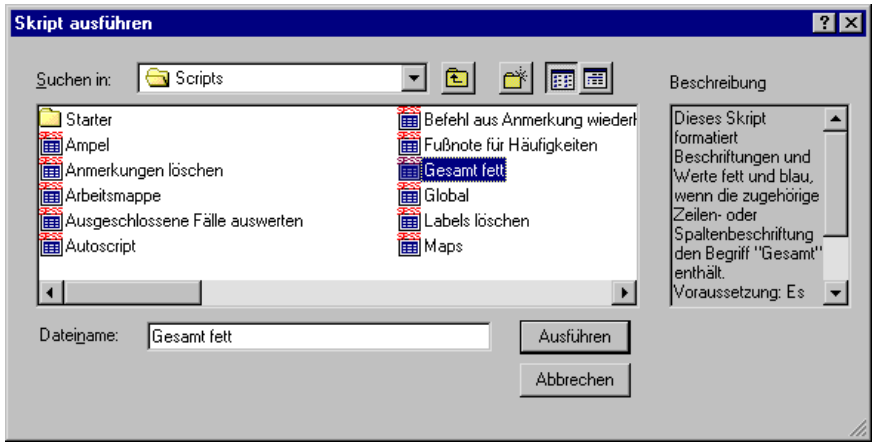
In SPSS verfügen Sie über mehrere Skripts inklusive Autoskripts, die automatisch ausgeführt werden, sobald ein bestimmter Ausgabetyp erstellt wird. Sie können die Skripts unverändert verwenden oder Ihren Wünschen anpassen. Falls Sie Ihre eigenen Skripts erstellen möchten, können Sie aus einer Anzahl von Starter-Skripts auswählen.

So führen Sie Skripts aus

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Extras
  Skript ausführen…

Abbildung 47-1
Dialogfeld "Skript ausführen"



- Wählen Sie den Ordner *Skripts* aus.
- Wählen Sie das gewünschte Skript aus.

Für weitere Informationen siehe "Anpassen von Menüs und Symbolleisten" in Kapitel 45 auf S. 687.

Im Lieferumfang von SPSS enthaltene Skripts

Die folgenden Skripts sind im Programm enthalten:

Ausgeschlossene Fälle auswerten. Wiederholt eine Faktoren- oder Diskriminanzanalyse mit in vorhergehenden Analysen nicht ausgewählten Fällen. Vor dem Ausführen des Skripts muss eine bei einer vorangegangenen Faktoren- oder Diskriminanzanalyse erstellte Anmerkungstabelle ausgewählt werden.

Sig in p ändern. Ändert in den Spaltenbeschriftungen aller Pivot-Tabellen *Sig.* zu *p=*. Die Tabelle muß vor Ausführung des Skripts ausgewählt werden.

Navigator aufräumen. Löscht alle Anmerkungstabellen aus einem Ausgabedokument. Das Dokument muß vor Ausführung des Skripts im Hauptfenster des Viewers geöffnet werden.

Fußnote für Häufigkeiten. Fügt die in einer Tabelle für die Häufigkeitsstatistik angezeigten Statistiken als Fußnoten in die entsprechende Häufigkeitstabelle für jede Variable ein. Die Tabelle für die Häufigkeitsstatistik muß vor Ausführung des Skripts ausgewählt werden.

Gesamt fett. Formatiert alle mit *Gesamt* beschrifteten Zeilen, Spalten oder Schichten von Daten einer Pivot-Tabelle fett und blau. Die Tabelle muß vor Ausführung des Skripts ausgewählt werden.

Mittelwert-Report. Extrahiert Informationen aus einer Tabelle von Mittelwerten und schreibt die Ergebnisse in mehrere ASCII-Dateien für die Ausgabe. Die Mittelwerttabelle muß vor Ausführung des Skripts ausgewählt werden.

Beschriftung entfernen. Löscht alle Zeilen- und Spaltenbeschriftungen aus der ausgewählten Pivot-Tabelle. Die Tabelle muß vor Ausführung des Skripts ausgewählt werden.

Befehl aus Anmerkung wiederholen. Führt den in der ausgewählten Anmerkungstabelle gefundenen Befehl unter Verwendung der aktiven Datendatei erneut aus. Falls keine Datendatei geöffnet ist, versucht das Skript, die ursprünglich verwendete Datendatei zu lesen. Die Anmerkungstabelle muß vor Ausführung des Skripts ausgewählt werden.

Maximales R-Quadrat. Formatiert in Zusammenfassungstabellen von Regressionsmodellen die Zeilen fett und blau, welche dem Modell entsprechen, in dem das korrigierte R^2 maximiert wird. Die Zusammenfassungstabelle muß vor Ausführung des Skripts ausgewählt werden.

Für weitere Informationen siehe “Optionen” in Kapitel 44 auf S. 665.

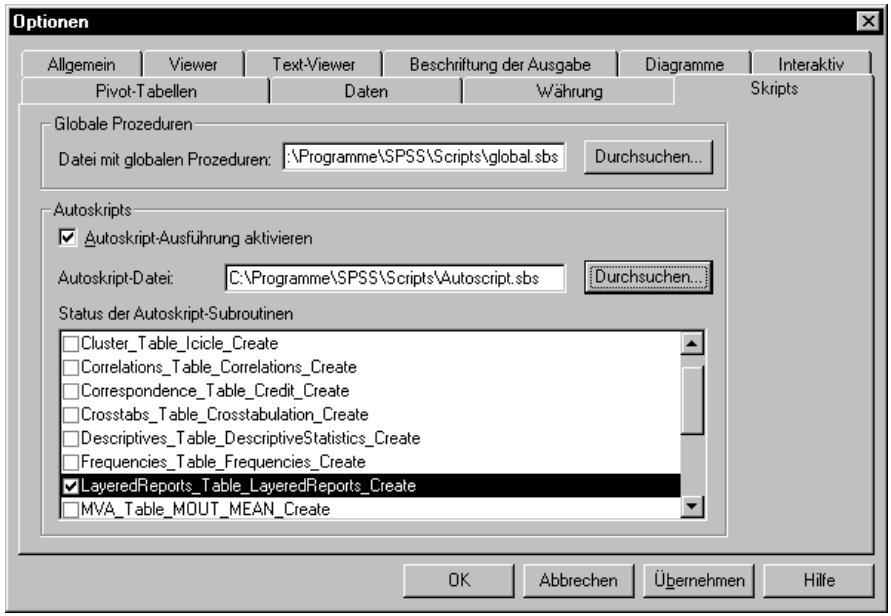
Anmerkung: Diese Liste ist möglicherweise nicht vollständig.

Autoskripts

Autoskripts werden automatisch ausgeführt, wenn sie durch die Erzeugung einer bestimmten Ausgabe durch eine gegebene Prozedur ausgelöst werden. SPSS enthält beispielsweise ein Autoskript, das automatisch die obere Diagonale entfernt und in der Tabelle die Korrelationskoeffizienten unterhalb einer bestimmten Signifikanz hervorhebt, sobald eine Korrelationstabelle mit der Prozedur “Bivariate Korrelationen” berechnet wurde.

Die auf Ihrem System zur Verfügung stehenden Autoskripts werden auf der Registerkarte “Skripts” (Menü “Bearbeiten”, Dialogfeld “Optionen”) angezeigt. Hier können Sie einzelne Skripts aktivieren oder deaktivieren.

Abbildung 47-2
Registerkarte “Skripts” im Dialogfeld “Optionen”



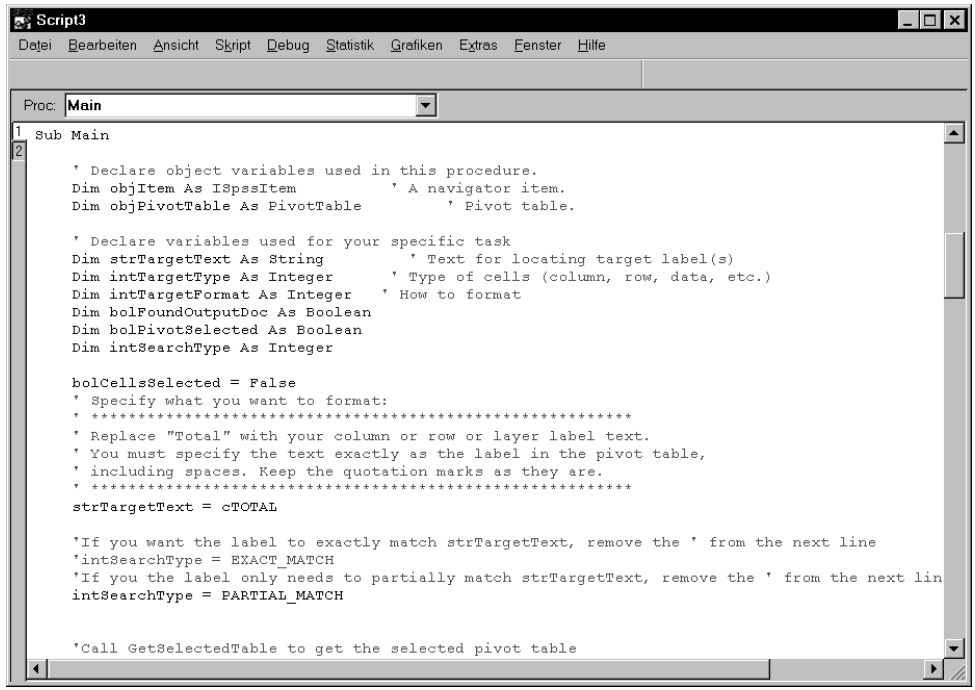
Autoskripts sind auf bestimmte Prozeduren und Ausgabetypen zugeschnitten. Ein Autoskript zum Formatieren von ANOVA-Tabellen, die mit einer einfaktoriellen ANOVA berechnet wurden, wird nicht durch ANOVA-Tabellen ausgelöst, die von anderen statistischen Prozeduren erstellt wurden. Sie können allerdings globale Prozeduren zum Erstellen weiterer Autoskripts für diese anderen ANOVA-Tabellen verwenden, die den Code weitgehend gemeinsam verwenden. Sie können aber auch für jeden Ausgabetyp einer Prozedur ein separates Autoskript verwenden. Mit “Häufigkeiten” werden zum Beispiel eine Häufigkeitstabelle und eine Statistiktabelle erstellt. Für jeden Tabellentyp können Sie ein separates Autoskript verwenden.

Für weitere Informationen siehe “Optionen” in Kapitel 44 auf S. 665.

Erstellen und Bearbeiten von Skripts

Sie können viele der mit der Software gelieferten Skripts an Ihre speziellen Zwecke anpassen. Es gibt beispielsweise ein Skript, das alle Anmerkungstabellen aus dem Viewer entfernt. Sie können dieses Skript leicht dahin gehend ändern, daß es jeden gewünschten Objekttyp und jedes gewünschte Label entfernt.

Abbildung 47-3
Ändern eines Skripts im Skript-Fenster


Script3
Datei Bearbeiten Ansicht Skript Debug Statistik Grafiken Extras Fenster Hilfe
Proc: Main
Sub Main
' Declare object variables used in this procedure.
Dim objItem As ISpssItem ' A navigator item.
Dim objPivotTable As PivotTable ' Pivot table.
' Declare variables used for your specific task
Dim strTargetText As String ' Text for locating target label(s)
Dim intTargetType As Integer ' Type of cells (column, row, data, etc.)
Dim intTargetFormat As Integer ' How to format
Dim bolFoundOutputDoc As Boolean
Dim bolPivotSelected As Boolean
Dim intSearchType As Integer
bolCellsSelected = False
' Specify what you want to format:
' **
' Replace "Total" with your column or row or layer label text.
' You must specify the text exactly as the label in the pivot table,
' including spaces. Keep the quotation marks as they are.
' **
strTargetText = cTOTAL
'If you want the label to exactly match strTargetText, remove the ' from the next line
'intSearchType = EXACT_MATCH
'If you the label only needs to partially match strTargetText, remove the ' from the next lin
intSearchType = PARTIAL_MATCH
'Call GetSelectedTable to get the selected pivot table

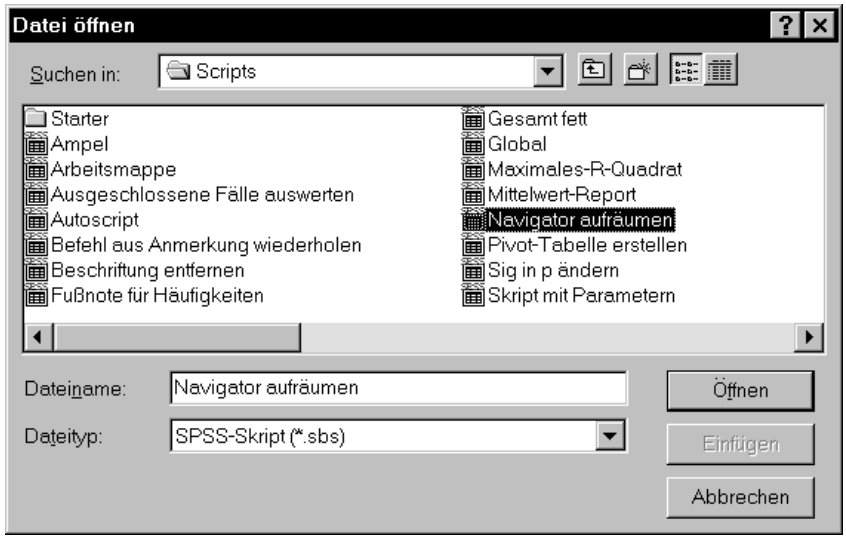
Falls Sie eigene Skripts erstellen möchten, können Sie aus einer Anzahl von Starter-Skripts auswählen.

So bearbeiten Sie ein Skript

- Wählen Sie die folgenden Befehle aus den Menüs aus:

Datei
  Öffnen
    Skript…

Abbildung 47-4
Öffnen einer Skript-Datei



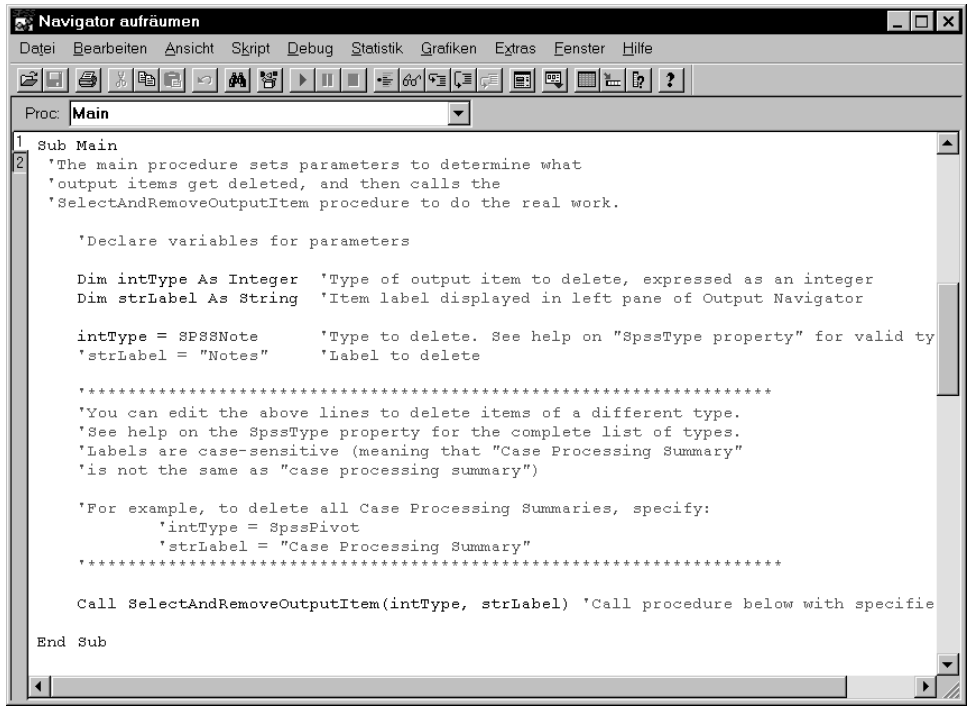
- Wählen Sie den Ordner *Skripts* aus.
- Wählen Sie den Dateityp SPSS-Skript (*.sbs) aus.
- Wählen Sie das gewünschte Skript aus.

  Falls Sie mehr als ein Skript auswählen, wird jedes in einem eigenen Fenster geöffnet.

Skript-Fenster

Das Skript-Fenster ist eine vollausgestattete Programmierumgebung für die Sprache “Sax BASIC” mit einem Dialogfeld-Editor, einem Objektkatalog, Debug-Funktionen und einer kontextsensitiven Hilfe.

Abbildung 47-5
Skript-Fenster


Navigator aufräumen
Datei Bearbeiten Ansicht Skript Debug Statistik Grafiken Extras Fenster Hilfe
Proc: Main
Sub Main
'The main procedure sets parameters to determine what
'output items get deleted, and then calls the
'SelectAndRemoveOutputItem procedure to do the real work.
'Declare variables for parameters
Dim intType As Integer 'Type of output item to delete, expressed as an integer
Dim strLabel As String 'Item label displayed in left pane of Output Navigator
intType = SPSSNote 'Type to delete. See help on "SpssType property" for valid ty
'strLabel = "Notes" 'Label to delete
'You can edit the above lines to delete items of a different type.
'See help on the SpssType property for the complete list of types.
'Labels are case-sensitive (meaning that "Case Processing Summary"
'is not the same as "case processing summary")
'For example, to delete all Case Processing Summaries, specify:
'intType = SpssPivot
'strLabel = "Case Processing Summary"
Call SelectAndRemoveOutputItem(intType, strLabel) 'Call procedure below with specifie
End Sub

- Wenn Sie den Cursor bewegen, wird der Name der aktuellen Prozedur am oberen Rand des Fensters angezeigt.
- Begriffe in blauer Farbe sind reservierte Wörter in BASIC, z. B. `Sub`, `End Sub` und `Dim`. Sie können kontextsensitive Hilfe zu diesen Begriffen aufrufen, indem Sie auf diese klicken und gleichzeitig F1 drücken.
- Bei violett formatierten Termen handelt es sich um Objekte, Eigenschaften oder Methoden. Sie können auch auf diese Begriffe klicken und gleichzeitig F1 drücken, um die Hilfe aufzurufen, jedoch nur, wenn diese Terme in gültigen Anweisungen stehen und violett formatiert sind. (Wenn Sie auf den Namen eines Objekts in einem Kommentar klicken, rufen Sie die Hilfe zur Sprache “Sax BASIC” und nicht die Hilfe zu SPSS-Objekten auf.)
- Kommentare werden grün angezeigt.
- Drücken Sie F2, um den Objektkatalog aufzurufen, mit dem Sie Objekte, Eigenschaften und Methoden anzeigen lassen können.

Eigenschaften des Skript-Editors (Skript-Fenster)

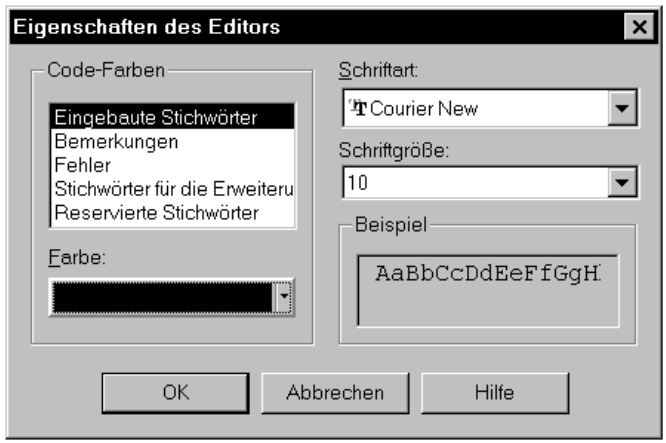
Im Skript-Fenster werden Code-Elemente zur besseren Unterscheidung farbig wiedergegeben. In der Standardeinstellung sind Kommentare grün, Sax BASIC-Begriffe blau und gültige Objekte, Eigenschaften und Methoden violett dargestellt. Sie können andere Farben für diese Elemente festlegen und Schriftgröße und Schriftart für den Text ändern.

So legen Sie die Eigenschaften des Skript-Editors fest:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Skript
  Eigenschaften des Editors…

  Abbildung 47-6
  Dialogfeld "Eigenschaften des Editors"

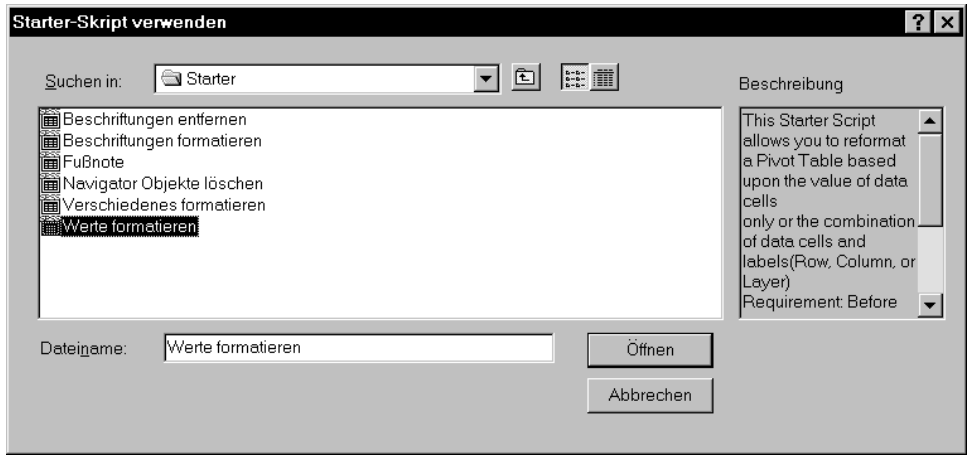


- Sie ändern die Farbe eines Code-Elementes, indem Sie das Element markieren und eine Farbe aus der Dropdown-Palette auswählen.

Starter-Skripts

Beim Erstellen eines neuen Skripts können Sie aus einer Anzahl von Starter-Skripts auswählen.

Abbildung 47-7
Dialogfeld "Starter-Skript verwenden"



Jedes Starter-Skript enthält den Code für mindestens eine der häufig verwendeten Prozeduren und Hinweise zum Anpassen des Skripts an Ihre speziellen Erfordernisse.

Beschriftungen entfernen. Löscht Zeilen oder Spalten in einer Pivot-Tabelle in Abhängigkeit vom Inhalt der Zeilen- oder Spaltenbeschriftung. Um dieses Skript ausführen zu können, muß im Dialogfeld "Tabelleneigenschaften" die Option Leere Zeilen und Spalten ausblenden aktiviert sein.

Navigator-Objekte löschen. Löscht Objekte anhand verschiedener Kriterien aus dem Viewer.

Fußnote. Formatiert die Fußnote einer Pivot-Tabelle neu, ändert den Text in einer Fußnote oder fügt eine Fußnote hinzu.

Beschriftungen formatieren. Formatiert anhand der Beschriftungen der Zeilen, Spalten oder Schichten eine Pivot-Tabelle neu.

Werte formatieren. Formatiert auf der Grundlage der Werte von Datenzellen oder einer Kombination aus Datenzellen und Beschriftungen Pivot-Tabellen neu.

Verschiedenes formatieren. Formatiert oder ändert den Text in der Kopfzeile einer Pivot-Tabelle, in einer Ecke oder in einer Erklärung.

Sie können auch jedes andere verfügbare Skript als Starter-Skript verwenden, das Anpassen kann bei diesen jedoch etwas komplizierter sein. Öffnen Sie das Skript und speichern Sie es unter einem anderen Dateinamen.

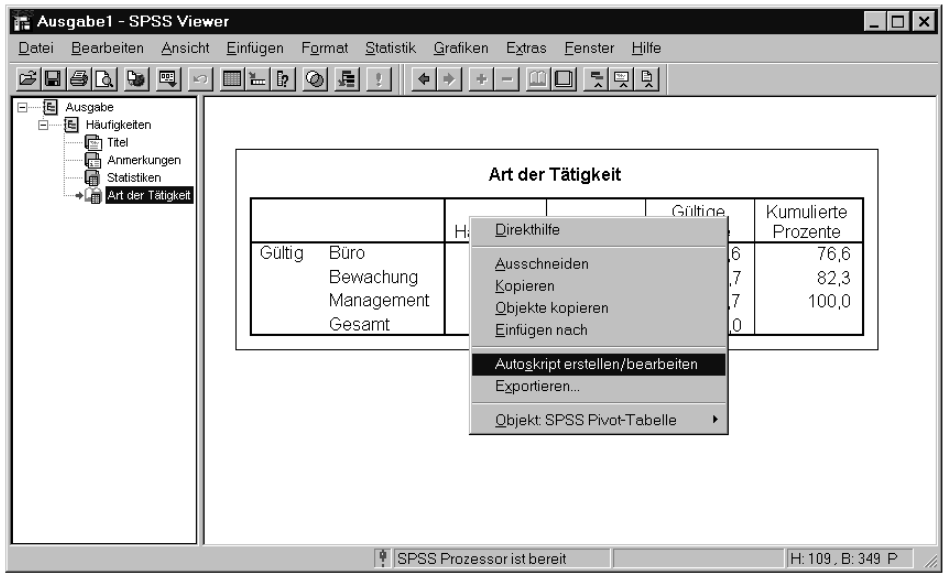
Erstellen von Skripts

- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Neu
  Skript…
- Wählen Sie gegebenenfalls ein Starter-Skript aus.
- Klicken Sie auf Abbrechen, wenn Sie kein Starter-Skript verwenden möchten.

Erstellen von Autoskripts

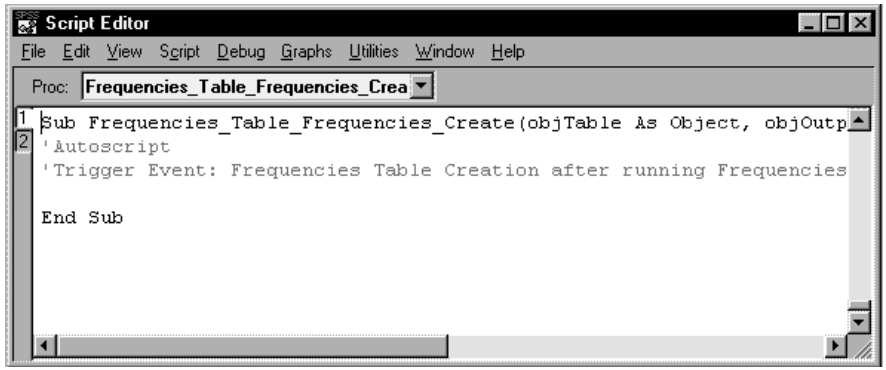
Beim Erstellen eines Autoskripts beginnen Sie mit dem Ausgabeobjekt, das als Auslöser für das Skript dienen soll. Wenn Sie beispielsweise ein Autoskript erstellen möchten, das bei jeder berechneten Häufigkeitstabelle ausgeführt wird, erstellen Sie wie gewohnt eine Häufigkeitstabelle und klicken im Viewer einmal auf die Tabelle, um diese auszuwählen. Sie können dann mit der rechten Maustaste klicken oder das Menü “Extras” verwenden, um ein neues Autoskript zu erstellen, das jedesmal aufgerufen wird, wenn eine Tabelle diesen Typs erzeugt wird.

Abbildung 47-8
Erstellen eines neuen Autoskripts



Standardmäßig wird jedes von Ihnen erstellte Autoskript als neue Prozedur zur aktuellen Autoskript-Datei (*autscript.sbs*) hinzugefügt. Der Name der Prozedur bezieht sich auf das Ereignis, welches die Prozedur aufruft. Wenn Sie beispielsweise ein Autoskript erstellen, das immer dann aufgerufen wird, wenn Sie mit der Prozedur “Explorative Datenanalyse” eine Tabelle für deskriptive Statistiken erstellen, trägt die Autoskript-Subroutine den Namen `Explore_Table_Descriptives_Create`.

Abbildung 47-9
Im Skript-Fenster angezeigte neue Autoskript-Prozedur


Script Editor
File Edit View Script Debug Graphs Utilities Window Help
Proc: Frequencies_Table_Frequencies_Crea
1 Sub Frequencies_Table_Frequencies_Create(objTable As Object, objOutp
2 'Autoscript
'Trigger Event: Frequencies Table Creation after running Frequencies
End Sub

Dies erleichtert das Entwickeln von Autoskripts, da Sie keinen Code schreiben müssen, um zu den zu bearbeitenden Objekten zu gelangen. Dies bedeutet aber auch, daß die Autoskripts jeweils nur bei einem bestimmten Ausgabetyp und einer bestimmten statistischen Prozedur funktionieren.

So erstellen Sie ein Autoskript

- ▶ Wählen Sie im Viewer das Objekt aus, durch welches das Autoskript ausgelöst werden soll.
- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

  Extras
  Autoskript erstellen/bearbeiten…

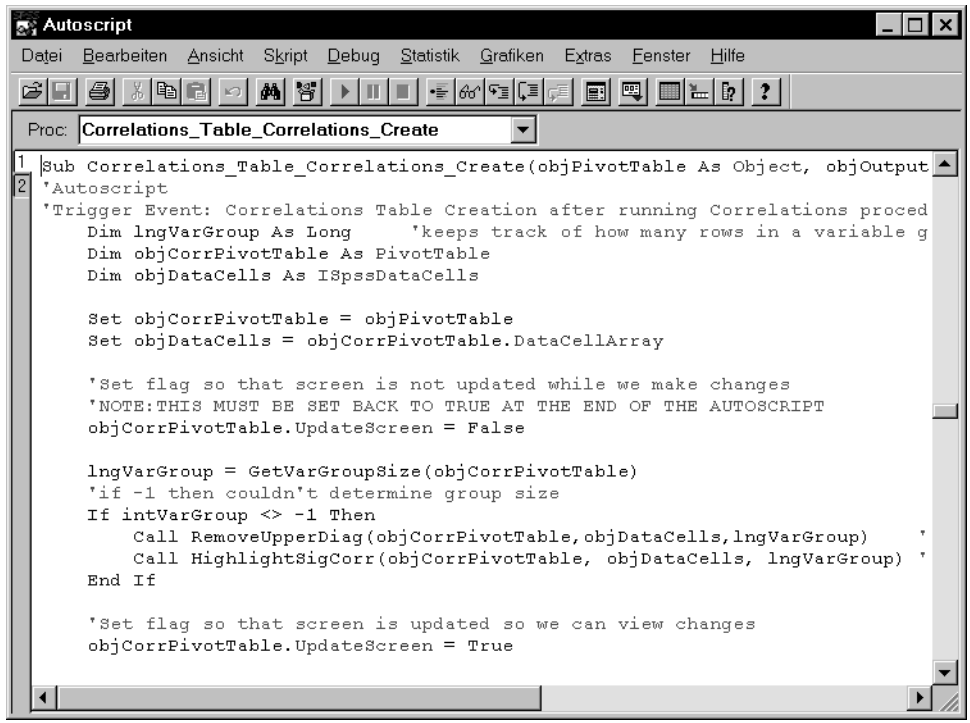
  Falls für das ausgewählte Objekt kein Autoskript vorhanden ist, wird ein neues Autoskript erstellt. Falls bereits ein Autoskript vorhanden ist, wird dieses angezeigt.
- ▶ Schreiben Sie den Code.
- ▶ Klicken Sie im Menü “Bearbeiten” auf Optionen, um das Autoskript zu aktivieren oder zu deaktivieren.

Ereignisse, die Autoskripts auslösen

Der Name der Autoskript-Prozedur bezieht sich auf das Ereignis, welches das Autoskript auslöst. Folgende Ereignisse können Autoskripts auslösen:

Erstellen von Pivot-Tabellen. Der Name der Prozedur bezieht sich auf den Tabellentyp und auf die Prozedur, mit der die Tabelle erstellt wurde, beispielweise `Correlations_Table_Correlations_Create`.

Abbildung 47-10
Autoskript-Prozedur für Korrelationstabellen


Autoscript
Datei Bearbeiten Ansicht Skript Debug Statistik Grafiken Extras Fenster Hilfe
Proc: Correlations_Table_Correlations_Create
Sub Correlations_Table_Correlations_Create(objPivotTable As Object, objOutput
'Autoscript
'Trigger Event: Correlations Table Creation after running Correlations proced
Dim lngVarGroup As Long 'keeps track of how many rows in a variable g
Dim objCorrPivotTable As PivotTable
Dim objDataCells As ISpssDataCells
Set objCorrPivotTable = objPivotTable
Set objDataCells = objCorrPivotTable.DataCellArray
'Set flag so that screen is not updated while we make changes
'NOTE:THIS MUST BE SET BACK TO TRUE AT THE END OF THE AUTOSCRIPT
objCorrPivotTable.UpdateScreen = False
lngVarGroup = GetVarGroupSize(objCorrPivotTable)
'if -1 then couldn't determine group size
If intVarGroup <> -1 Then
Call RemoveUpperDiag(objCorrPivotTable,objDataCells,lngVarGroup)
Call HighlightSigCorr(objCorrPivotTable, objDataCells, lngVarGroup)
End If
'Set flag so that screen is updated so we can view changes
objCorrPivotTable.UpdateScreen = True

Erstellen von Titeln. Bezogen auf die zum Erstellen verwendete statistische Prozedur: `Correlations_Title_Create`.

Erstellen von Anmerkungen. Bezogen auf die zum Erstellen verwendete Prozedur: `Correlations_Notes_Create`.

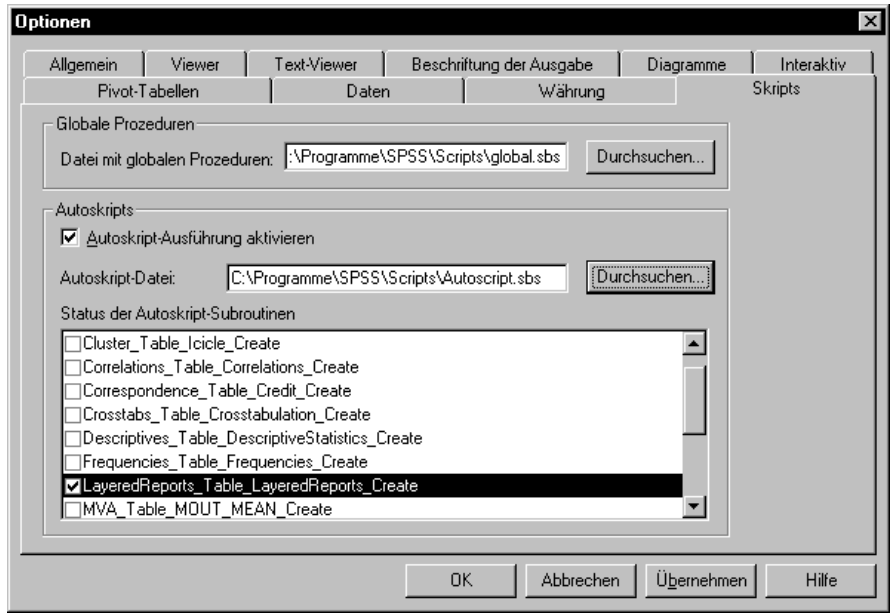
Erstellen von Warnungen. Bezogen auf die zum Erstellen verwendete Prozedur.

Sie können Skripts auch verwenden, um Autoskripts indirekt auszulösen. Sie können z. B. ein Skript schreiben, das die Prozedur "Korrelationen" aufruft, die wiederum ein Autoskript auslöst, das mit der resultierenden Korrelationstabelle verknüpft ist.

Autoskript-Datei

Alle Autoskripts werden in einer einzigen Datei gespeichert. Andere Skripts werden dagegen jeweils in einer eigenen Datei gespeichert. Jedes neu erstellte Autoskript wird dieser Datei hinzugefügt. Der Name der aktuellen Autoskript-Datei wird auf der Registerkarte "Skripts" (Menü "Bearbeiten", Dialogfeld "Optionen") angezeigt.

Abbildung 47-11
Im Dialogfeld "Optionen" angezeigte Autoskript-Subroutinen



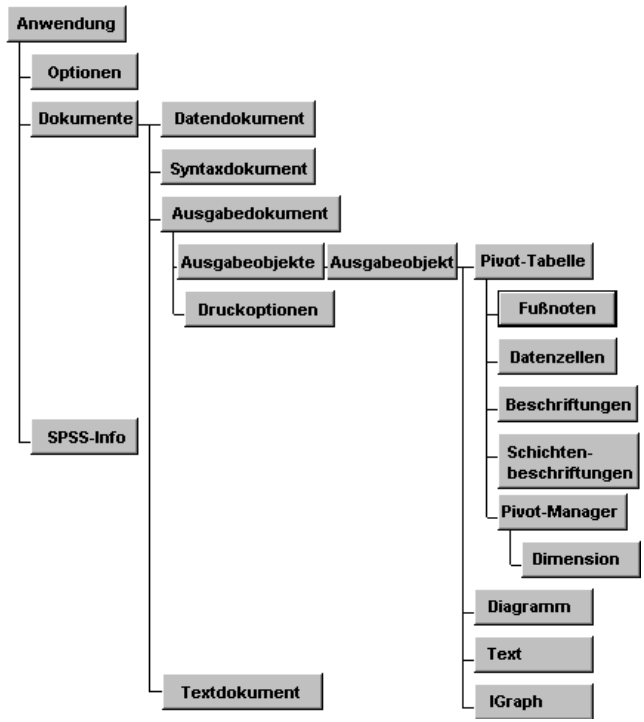
Im Dialogfeld "Optionen" werden alle Autoskripts der gegenwärtig ausgewählten Autoskript-Datei angezeigt, so daß Sie einzelne Skripts aktivieren und deaktivieren können.

Die Standard-Autoskript-Datei ist *autscript.sbs*. Sie können auch eine andere Autoskript-Datei angeben, aber es kann jeweils nur eine aktiv sein.

Funktionsweise von Skripts

Durch Skripts werden Objekte unter Verwendung von Eigenschaften und Methoden geändert. Pivot-Tabellen sind beispielsweise eine Klasse von Objekten. Bei Objekten dieser Klasse können Sie die Methode “SelectTable” verwenden, um alle Elemente in der Tabelle auszuwählen, und die Eigenschaft “TextColor”, um die Farbe das ausgewählten Textes zu ändern. Zu jeder Objektklasse gehören bestimmte Eigenschaften und Methoden. Alle Klassen bzw. SPSS-Typen von SPSS-Objekten sind in der SPSS-Typenbibliothek zusammengefaßt.

Abbildung 47-12
Baumansicht der Objekte



Die Verwendung von Objekten erfolgt in zwei Schritten. Zuerst stellen Sie einen Verweis auf das Objekt her. Dies wird als *Zuweisen* des Objekts bezeichnet. Dann verwenden Sie Eigenschaften und Methoden, um etwas zu ändern. Sie können Objekte zuweisen, indem Sie durch die Hierarchie der Objekte navigieren. Bei jedem

Schritt verwenden Sie Eigenschaften oder Methoden von Objekten, die höher in der Hierarchie angeordnet sind, um die darunterliegenden Objekte zuzuweisen. Wenn Sie beispielsweise ein Pivot-Tabellen-Objekt zuweisen möchten, müssen Sie zuerst den Viewer zuweisen, der die Pivot-Tabelle enthält, und dann die Objekte im Viewer.

Jedes Objekt, das Sie zuweisen, wird in einer Variablen gespeichert. Beachten Sie, daß der tatsächliche Inhalt der Variablen stets nur ein Verweis auf das Objekt ist. Einer der ersten Schritte beim Erstellen von Skripts ist oft das Deklarieren von Variablen für die von Ihnen benötigten Objekte.

Tipp: Ohne eine Vorstellung von der Funktionsweise des Programms ist es schwierig, die Funktionsweise von Skripts zu verstehen. Bevor Sie ein Skript schreiben, führen Sie die Aktion einige Male wie gewohnt mit der Maus aus. Beachten Sie bei jedem Schritt, welche Objekte Sie bearbeiten und welche Eigenschaften eines jeden Objekts Sie ändern.

Deklarieren von Variablen (Skripts)

Es empfiehlt sich, alle Variablen vor ihrer Verwendung zu deklarieren, auch wenn dies nicht immer erforderlich ist. Dies geschieht meistens unter Verwendung der Deklarationsanweisung `Dim`:

```
Dim objOutputDoc As ISpssOutputDoc
Dim objPivotTable As PivotTable
Dim intType As Integer
Dim strLabel As String
```

Mit jeder Deklaration werden Name und Typ der Variablen festgelegt. Mit der ersten oben aufgeführten Deklaration wird beispielsweise eine Objektvariable mit dem Namen `objOutputDoc` erstellt und der Objektklasse `ISpssOutputDoc` zugewiesen. Die Variable hat noch keinen Wert, da sie noch keinem bestimmten Viewer zugewiesen wurde. Die Anweisung deklariert also lediglich das Vorhandensein der Variablen. (Dieser Prozeß wurde als “Umbenennen der Objekte, die Sie verwenden möchten” bezeichnet.)

Konventionen für die Benennung von Variablen. Gemäß der Konvention gibt der Variablenname den Variablentyp an. Namen für Objektvariablen beginnen mit `obj`, ganzzahlige Variablen beginnen mit `int`, und String-Variablen beginnen mit `str`.

Dies sind lediglich Konventionen. Sie können Variablen nach Wunsch benennen. Diese Konventionen erleichtern Ihnen jedoch erheblich das Verstehen ihres Codes.

SPSS-Objektklassen. `ISpssOutputDoc` und `PivotTable` sind Namen von SPSS-Objektklassen. Jede Klasse steht für einen Objekttyp, der mit dem Programm erstellt werden kann, wie zum Beispiel ein Viewer oder eine Pivot-Tabelle. Zu jeder Objektklasse gehören bestimmte Eigenschaften und Methoden. Die Sammlung aller SPSS-Objektklassen, auch SPSS-Typen genannt, wird als SPSS-Typenbibliothek bezeichnet.

SPSS-Objektklassen und Konventionen für die Benennung von Variablen

Die folgenden Variablennamen werden in den mit dem Programm gelieferten Beispielskripts verwendet. Es empfiehlt sich, diese bei allen Skripts zu verwenden. Beachten Sie, daß die Namen aller Objektklassen mit Ausnahme von Pivot-Tabellen mit `ISpss` beginnen.

Objekt	Typ oder Klasse	Variablenname
SPSS-Anwendung	`IspssApp`	`objSpssApp` ist eine globale Variable und muß nicht deklariert werden.
SPSS-Optionen	`ISpssOptions`	`objSpssOptions`
SPSS-Datei-Information	`ISpssInfo`	`objSpssInfo`
Dokumente	`ISpssDocuments`	`objDocuments`
Datendokument	`ISpssDataDoc`	`objDataDoc`
Syntax-Dokument	`ISpssSyntaxDoc`	`objSyntaxDoc`
Viewer-Dokument	`ISpssOutputDoc`	`objOutputDoc`
Druckoptionen	`ISpssPrintOptions`	`objPrintOptions`
Sammlung von Ausgabeobjekten	`ISpssItems`	`objOutputItems`
Ausgabeobjekt	`ISpssItem`	`objOutputItem`
Diagramme	`ISpssChart`	`objSPSSChart`
Text	`ISpssRtf`	`objSPSSText`
Pivot-Tabelle	`PivotTable`	`objPivotTable`
Fußnoten	`ISpssFootnotes`	`objFootnotes`

Objekt	Typ oder Klasse	Variablenname
Datenzellen	`ISpssDataCells`	`objDataCells`
Schichtenbeschriftungen	`ISpssLayerLabels`	`objLayerLabels`
Spaltenbeschriftungen	`ISpssLabels`	`objColumnLabels`
Zeilenbeschriftungen	`ISpssLabels`	`objRowLabels`
Pivot-Manager	`ISpssPivotMgr`	`objPivotMgr`
Dimension	`ISpssDimension`	`objDimension`

Zuweisen von Automatisierungsobjekten (Skripts)

Ein Objekt *zuzuweisen* bedeutet, einen Verweis auf ein Objekt herzustellen. So können Sie Eigenschaften und Methoden verwenden, um eine Aktion vorzunehmen. Jeder Objektverweis wird in einer Variablen gespeichert. Sie weisen ein Objekt zu, indem Sie zuerst eine Objektvariable der geeigneten Klasse deklarieren und dann die Variable auf das bestimmte Objekt setzen. So weisen Sie beispielsweise den vorgegebenen Viewer zu:

```
Dim objOutputDoc As ISpssOutputDoc
Set objOutputDoc = objSpssApp.GetDesignatedOutputDoc
```

Eigenschaften und Methoden von Objekten, die in der Objekthierarchie höher stehen, werden zum Zuweisen der darunter befindlichen Objekte verwendet. Die zweite Anweisung oben weist das Haupt-Ausgabedokument mit `GetDesignatedOutputDoc` zu. Dies ist eine Methode, die mit dem Anwendungsobjekt (dem Objekt auf höchster Ebene) verbunden ist. In ähnlicher Weise werden Pivot-Tabellen-Objekte zugewiesen. Sie weisen zuerst den Viewer zu, der die Pivot-Tabelle enthält, dann die Objektgruppen in diesem Viewer und so weiter.

Beispiel: Zuweisen eines Ausgabeobjekts

Dieses Skript weist das dritte Ausgabeobjekt im vorgegebenen Viewer zu und aktiviert es. Wenn dieses Objekt kein OLE-Objekt ist, erzeugt dieses Skript eine Fehlermeldung.

```
Sub Main
```

```
Dim objOutputDoc As ISpssOutputDoc'Objektvariablen deklarieren
Dim objOutputItems As ISpssItems
Dim objOutputItem As ISpssItem

Set objOutputDoc = objSpssApp.GetDesignatedOutputDoc'Verweis auf Haupt-Ausgabedokument zuweisen
Set objOutputItems = objOutputDoc.Items() 'Sammlung der Objekte im Dokument zuweisen
Set objOutputItem = objOutputItems.GetItem(2) 'Drittes Ausgabeobjekt zuweisen
'(Objektnummern beginnen bei 0, so daß mit "2" das dritte zugewiesen wird)

objOutputItem.Activate 'Ausgabeobjekt aktivieren

End sub
```

Beispiel: Zuweisen der ersten Pivot-Tabelle

Dieses Skript weist die erste Pivot-Tabelle im vorgegebenen Viewer zu und aktiviert diese.

```
Sub Main

Dim objOutputDoc As ISpssOutputDoc'Objektvariablen deklarieren
Dim objOutputItems As ISpssItems
Dim objOutputItem As ISpssItem
Dim objPivotTable As PivotTable

Set objOutputDoc = objSpssApp.GetDesignatedOutputDoc'Verweis auf Haupt-Ausgabedokument zuweisen
Set objOutputItems = objOutputDoc.Items()'Sammlung der Objekte im Dokument zuweisen

Dim intItemCount As Integer'Anzahl von Ausgabeobjekten
Dim intItemType As Integer'Objekttyp (definiert durch Eigenschaft "SpssType")

intItemCount = objOutputItems.Count()'Anzahl von Ausgabeobjekten zuweisen
   For index = 0 To intItemCount'Ausgabeobjekte durchlaufen
   Set objOutputItem = objOutputItems.GetItem(index)'Aktuelles Objekt zuweisen
      intItemType = objOutputItem.SPSSType()'Typ des aktuellen Objekts zuweisen
      If intItemType = SPSSPivot Then
         Set objPivotTable = objOutputItem.Activate()'Wenn Objekt eine Pivot-Tabelle ist, aktivieren
         Exit For
      End If
Next index
```

```
End sub
```

Diese Beispiele sind auch über die Online-Hilfe verfügbar. Sie können diese selbst ausprobieren, indem Sie den Code aus der Hilfe in das Skript-Fenster einfügen.

Eigenschaften und Methoden (Skripts)

Wie reale Objekte besitzen OLE-Automatisierungsobjekte Merkmale und Funktionen. In der Programmierterminologie werden die Merkmale als Eigenschaften und die Funktionen als Methoden bezeichnet. Jede Objektklasse verfügt über bestimmte Eigenschaften und Methoden, welche die Verwendungsmöglichkeiten dieser Objekte bestimmen.

Objekt	Eigenschaft	Methode
Bleistift (reales Objekt)	Härte Farbe	Schreiben Radieren
Pivot-Tabelle (SPSS)	TextFont DataCellWidths CaptionText	SelectTable ClearSelection HideFootnotes

Beispiel: Verwenden von Eigenschaften (Skripts)

Mit Eigenschaften können die Parameter von Objekten eingestellt oder zurückgegeben werden, wie die Farbe oder die Zellenbreite. Wenn eine Eigenschaft auf der linken Seite des Gleichheitszeichens steht, können Sie diese bearbeiten. Wenn Sie also die Erklärung für eine aktivierte Pivot-Tabelle (`objPivotTable`) auf `Anitas Ergebnisse` setzen möchten, schreiben Sie:

```
objPivotTable.CaptionText = "Anitas Ergebnisse"
```

Wenn eine Eigenschaft rechts vom Gleichheitszeichen steht, können Sie deren Inhalt lesen. So erhalten Sie beispielsweise eine Erklärung der aktivierten Pivot-Tabelle und speichern diese in einer Variablen:

```
strFontName = objPivotTable.CaptionText
```

Beispiel: Verwenden von Methoden (Skripts)

Mit Methoden können Sie Aktionen an Objekten vornehmen. Sie können z. B. alle Elemente in einer Tabelle auswählen:

```
objPivotTable.SelectTable
```

Sie können eine Auswahl auch rückgängig machen:

```
objPivotTable.ClearSelection
```

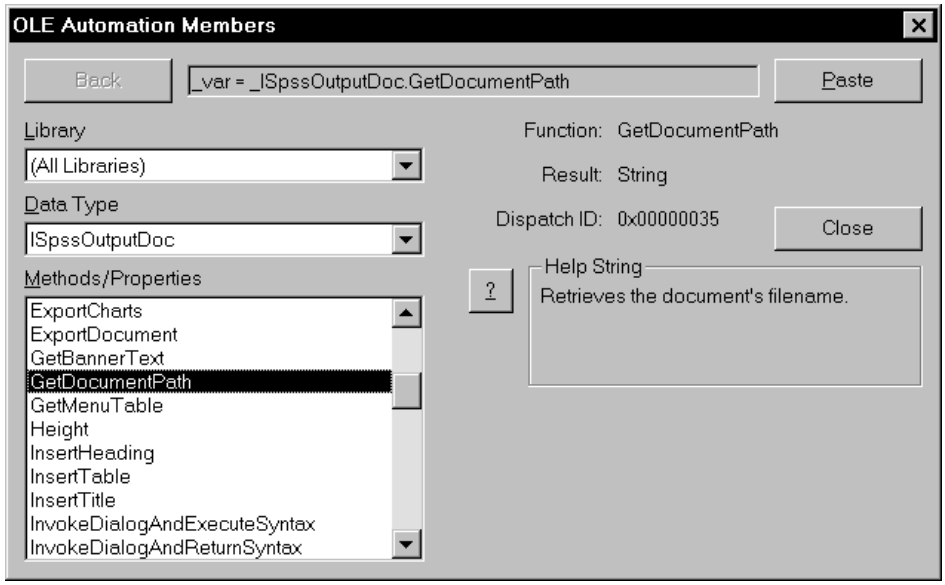
Einige Methoden geben andere Objekte zurück. Diese Methoden sind zum Navigieren durch die Objekthierarchie besonders wichtig. Die Methode `GetDesignatedOutputDoc` gibt beispielsweise das Haupt-Ausgabedokument zurück. Dies ermöglicht Ihnen den Zugriff auf die Objekte in diesem Ausgabedokument:

```
Set objOutputDoc = objSpssApp.GetDesignatedOutputDoc
Set objItems = objOutputDoc.Items
```

Objektkatalog

Im Objektkatalog werden alle Objektklassen und die mit ihnen verbundenen Methoden und Eigenschaften angezeigt. Sie können auch auf Hilfefunktionen zu einzelnen Eigenschaften und Methoden zugreifen und die ausgewählten Eigenschaften und Methoden in Ihr Skript einfügen.

Abbildung 47-13
Objektkatalog



Arbeiten mit dem Objektkatalog

- Wählen Sie die folgenden Befehle aus den Menüs des Skript-Fensters aus:

  Debug
  Objektkatalog…

- Wählen Sie aus der Liste der Datentypen eine Objektklasse aus, um sich Methoden und Eigenschaften für diese Klasse anzeigen zu lassen.

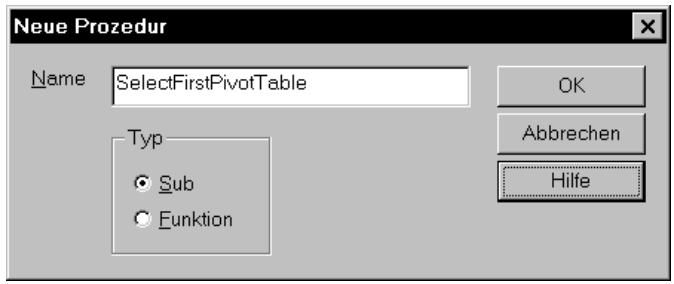
- Wählen Sie Eigenschaften und Methoden aus, um die kontextsensitive Hilfe aufzurufen oder um diese in Ihr Skript einzufügen.

Neue Prozeduren (Skripts)

Eine Prozedur ist eine mit einem Namen versehene Folge von Anweisungen, die als Block ausgeführt werden. Das Organisieren von Befehlsfolgen in Prozeduren erleichtert die Bearbeitung und Wiederverwendung von Programmabschnitten. Skripts müssen mindestens eine Prozedur (die Subroutine `Main`) enthalten. Meistens

bestehen sie jedoch aus mehreren Prozeduren. Die Prozedur `Main` enthält neben Aufrufen von Subroutinen, mit denen die eigentliche Aufgabe abgearbeitet wird, möglicherweise nur wenige Befehle.

Abbildung 47-14
Dialogfeld "Neue Prozedur"



Prozeduren können Subroutinen oder Funktionen sein. Eine Prozedur beginnt mit einer Anweisung, die den Typ der Prozedur und den Namen festlegt (beispielsweise `Sub Main` oder `Function DialogMonitor( )`), und endet mit der passenden End-Anweisung (`End Sub` oder `End Function`).

Wenn Sie einen Bildlauf durch das Skript-Fenster durchführen, wird der Name der aktuellen Prozedur am oberen Rand des Skript-Fensters angezeigt. Sie können jede Prozedur in einem Skript beliebig oft aufrufen. Sie können außerdem jede Prozedur aus der globalen Skriptdatei aufrufen, was die Mehrfachverwendung von Prozeduren in verschiedenen Skripts erleichtert.

So fügen Sie einem Skript eine neue Prozedur hinzu

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Skript
  Neue Prozedur...

- Geben Sie einen Namen für die Prozedur ein.

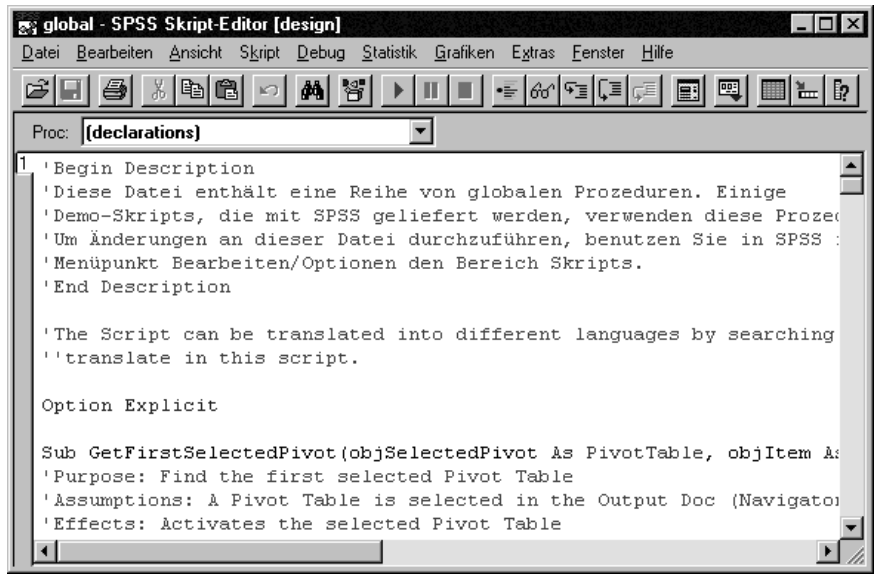
- Aktivieren Sie das Optionsfeld Sub oder Funktion.

Wahlweise können Sie eine neue Prozedur erstellen, indem Sie die Anweisungen, mit denen die Prozedur definiert wird, direkt in das Skript schreiben.

Globale Prozeduren (Skripts)

Falls Sie eine Prozedur oder Funktion in mehreren verschiedenen Skripts verwenden möchten, können Sie diese der globalen Skriptdatei hinzufügen. Prozeduren aus der globalen Skriptdatei können von allen anderen Skripts aufgerufen werden.

Abbildung 47-15
Globale Skriptdatei


global - SPSS Skript-Editor [design]
Datei Bearbeiten Ansicht Skript Debug Statistik Grafiken Extras Fenster Hilfe
Proc: (declarations)
'Begin Description
'Diese Datei enthält eine Reihe von globalen Prozeduren. Einige
'Demo-Skripts, die mit SPSS geliefert werden, verwenden diese Proze
'Um Änderungen an dieser Datei durchzuführen, benutzen Sie in SPSS
'Menüpunkt Bearbeiten/Optionen den Bereich Skripts.
'End Description
'The Script can be translated into different languages by searching
''translate in this script.
Option Explicit
Sub GetFirstSelectedPivot(objSelectedPivot As PivotTable, objItem As
'Purpose: Find the first selected Pivot Table
'Assumptions: A Pivot Table is selected in the Output Doc (Navigato
'Effects: Activates the selected Pivot Table

Standardmäßig wird als globales Skript *global.sbs* verwendet. Sie können nach Belieben Prozeduren zu dieser Datei hinzufügen. Auf der Registerkarte “Skripts” im Dialogfeld “Optionen” (Menü “Bearbeiten”) können Sie eine andere globale Datei festlegen. Es kann aber nur eine Datei auf einmal als globale Datei aktiviert sein. Dies bedeutet, daß die Prozeduren und Funktionen aus *global.sbs* nicht mehr verfügbar sind, wenn Sie eine neue globale Datei erstellen und diese als aktuelle globale Datei festlegen.

Sie können sich die globale Skriptdatei in jedem Skript-Fenster anzeigen lassen, indem Sie unterhalb der Symbolleiste auf der linken Seite des Fensters auf die Registerkarte “2” klicken. Sie können diese jedoch nur in einem Fenster gleichzeitig bearbeiten.

Globale Prozeduren müssen von anderen Skript-Prozeduren aus aufgerufen werden. Sie können ein globales Skript nicht direkt aus dem Menü “Extras” oder einem Skript-Fenster ausführen.

Hinzufügen einer Beschreibung zu einem Skript

Sie können eine Beschreibung hinzufügen und diese in den Dialogfeldern “Skript ausführen” und “Starter-Skript verwenden” anzeigen lassen. Sie beginnen den Kommentar mit der Zeile `Begin Description`, schreiben den gewünschten Kommentar, der eine oder mehrere Zeilen lang sein kann, und beenden ihn mit der Zeile `End Description`. Ein Beispiel:

```
'Begin Description
'Ändert in den Spaltenbeschriftungen aller Pivot-Tabellen "Sig." zu "p=".
'Anforderung: Die Pivot-Tabelle, die Sie ändern möchten, muß ausgewählt sein.
'End Description
```

Die Beschreibung muß als Kommentar formatiert sein, das heißt, jede Zeile muß mit einem Apostroph beginnen.

Skripts mit benutzerdefinierten Dialogfeldern

Ein benutzerdefiniertes Dialogfeld wird in zwei Schritten erzeugt: Zuerst erstellen Sie das Dialogfeld mit dem Dialog-Editor. Anschließend erstellen Sie die Funktion `DialogFunc`, mit welcher das Dialogfeld überwacht und sein Verhalten definiert wird.

Das Dialogfeld selbst wird durch den Block `Begin Dialog...End Dialog` definiert. Sie müssen den Code nicht direkt eingeben. Mit dem Dialog-Editor können Sie Dialogfelder grafisch entwerfen.

Abbildung 47-16
Erstellen eines Dialogfelds im Dialog-Editor

Der Editor zeigt zunächst ein leeres Dialogfeld an. Sie können Steuerelemente wie Schaltflächen und Kontrollkästchen hinzufügen, indem Sie das geeignete Symbol auswählen und dieses mit der Maus verschieben. Zeigen Sie mit der Maus auf ein Symbol, um eine Beschreibung des Symbols aufzurufen. Sie können auch die Seiten und Ecken des Dialogfelds ziehen, um dessen Größe zu ändern. Nachdem Sie ein Steuerelement hinzugefügt haben, klicken Sie mit der rechten Maustaste auf das Element, um dessen Eigenschaften festzulegen.

Überwachungsfunktion für das Dialogfeld. Sie legen eine Überwachungsfunktion fest, indem Sie mit der rechten Maustaste auf das Dialogfeld klicken, wobei kein Steuerelement in dem Feld ausgewählt sein darf. Geben Sie dann den Namen für die Funktion in das Feld “Dialog Function” ein. Die Anweisungen, welche die Funktionen definieren, werden Ihrem Skript hinzugefügt. Sie müssen die Funktion allerdings manuell bearbeiten, um deren Verhalten bei jeder Aktion festzulegen.

Klicken Sie anschließend auf die Schaltfläche “Save and Exit”, die Sie ganz rechts auf der Symbolleiste finden, um Ihrem Skript den Code für das Dialogfeld hinzuzufügen.

So erstellen Sie ein benutzerdefiniertes Dialogfeld

- Klicken Sie im Skript-Fenster mit dem Mauszeiger auf die Stelle des Skripts, an der Sie den Code für das Dialogfeld einfügen möchten.
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Skript
  Dialog-Editor…
- Wählen Sie Symbole aus der Palette aus, und ziehen Sie diese in das neue Dialogfeld, um Steuerelemente wie z. B. Schaltflächen oder Kontrollkästchen hinzuzufügen.
- Ändern Sie die Größe des Dialogfelds durch Ziehen der Seiten und Ecken.
- Klicken Sie mit der rechten Maustaste auf das Dialogfeld, wobei kein Steuerelement ausgewählt sein darf. Geben Sie in das Feld "Dialog Function" einen Namen für die Überwachungsfunktion des Dialogfelds ein.
- Klicken Sie anschließend auf die Schaltfläche "Save and Exit", die Sie ganz rechts auf der Symbolleiste finden.

  Sie müssen die Überwachungsfunktion für das Dialogfeld manuell bearbeiten, um das Verhalten des Dialogfelds festzulegen.

Überwachungsfunktionen für Dialogfelder (Skripts)

Die Überwachungsfunktion legt das Verhalten eines Dialogfelds für jeweils eine Anzahl angegebener Fälle fest. Die Funktion zeigt folgendes Grundverhalten:

```
Function DialogFunc(strDlgItem as String, intAction as Integer, intSuppValue as Integer)
  Select Case intAction
    Case 1 ' Initialisierung des Dialogfelds
    ...   'Anweisungen, die bei Initialisierung des Dialogfelds ausgeführt werden sollen
    Case 2 ' Wertänderung oder Klicken auf Schaltfläche
    ...   'Anweisungen...
    Case 3 'Text von TextBox oder ComboBox geändert ...
    Case 4 ' Fokus geändert ...
    Case 5 ' Leerlauf ...
  End Select
End Function
```

Parameter. Die Funktion muß in der Lage sein, drei Parameter zu übergeben: einen String (`strDlgItem`) und zwei Ganzzahlen (`intAction` und `intSuppValue`). Die Parameter sind Werte, die in Abhängigkeit von der jeweils durchgeführten Aktion von der Funktion an das Dialogfeld übergeben werden.

Wenn Sie im Dialogfeld beispielsweise auf ein Steuerelement klicken, wird der Name des Steuerelements an die Funktion als `strDlgItem` übergeben. Der Feldname wird in der Definition des Dialogfelds festgelegt. Der zweite Parameter (`intAction`) ist ein numerischer Wert, der angibt, welcher Vorgang im Dialogfeld stattgefunden hat. Der dritte Parameter wird in einigen Fällen für zusätzliche Informationen verwendet. Sie müssen alle drei Parameter in der Definition Ihrer Funktion aufführen, auch wenn Sie nicht alle verwenden.

Select Case intAction. Der Wert von `intAction` gibt an, welcher Vorgang im Dialogfeld stattgefunden hat. Wenn das Dialogfeld beispielsweise initialisiert wird, ist `intAction` = 1, wenn der Benutzer auf eine Schaltfläche klickt, nimmt `intAction` den Wert 2 an usw. Es sind fünf Aktionen möglich. Sie können Anweisungen festlegen, die wie im folgenden aufgeführt bei den Aktionen ausgeführt werden. Sie brauchen nicht alle fünf möglichen Fälle festzulegen, sondern nur die jeweils zutreffenden. Wenn Sie zum Beispiel bei der Initialisierung keine Anweisungen ausgeführt haben möchten, lassen Sie Fall 1 aus.

- **Case intAction = 1.** Legen Sie Anweisungen fest, die bei der Initialisierung des Dialogfelds ausgeführt werden sollen. Sie können beispielsweise eines oder mehrere Steuerelemente deaktivieren oder einen Signalton hinzufügen. Der String `strDlgItem` ist ein leerer String; `intSuppValue` ist gleich 0.
- **Case 2.** Wird ausgeführt, wenn auf eine Schaltfläche geklickt wird oder sich ein Wert in einem Steuerelement vom Typ `CheckBox`, `DropListBox`, `ListBox` oder `OptionGroup` ändert. Wenn auf eine Schaltfläche geklickt wird, stellt `strDlgItem` die Schaltfläche dar, `intSuppValue` ist bedeutungslos, und Sie müssen `DialogFunc = True` setzen, damit das Dialogfeld nicht geschlossen wird. Wenn ein Wert geändert wird, stellt `strDlgItem` das Objekt dar, dessen Wert sich geändert hat, und `intSuppValue` entspricht dem neuen Wert.
- **Case 3.** Wird ausgeführt, wenn sich ein Wert in einem Steuerelement vom Typ `TextBox` oder `ComboBox` ändert. Der String `strDlgItem` stellt das Steuerelement dar, dessen Text geändert wird und das den Fokus verliert. `intSuppValue` entspricht der Anzahl der Zeichen.

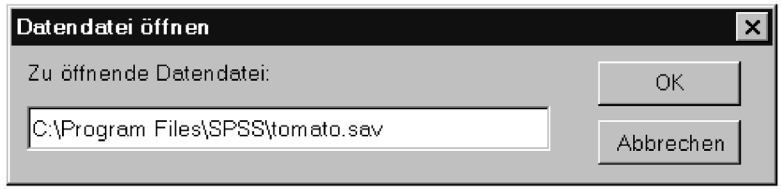
- **Case 4.** Wird ausgeführt, wenn sich der Fokus in dem Dialogfeld ändert. Der String `strDlgItem` erhält den Fokus, und `intSuppValue` ist das Objekt, das den Fokus verliert (das erste Objekt ist gleich 0, das zweite gleich 1 usw.).
- **Case 5.** Leerlauf. Der String `strDlgItem` ist ein leerer String, und `intSuppValue` ist gleich 0. Setzen Sie `DialogFunc = True`, um weiterhin Leerlaufaktionen zu empfangen.

Weitere Informationen hierzu erhalten Sie, indem Sie in der Hilfedatei "Sax BASIC Language Reference" unter den Beispielen und dem Prototyp `DialogFunc` nachschlagen.

Beispiel: Erstellen eines Skripts für ein einfaches Dialogfeld

Mit diesem Skript erstellen Sie ein einfaches Dialogfeld zum Öffnen einer Datendatei. Erläuterungen zur Subroutine `BuildDialog` und zur Überwachungsfunktion für das Dialogfeld erhalten Sie in den zugehörigen Abschnitten.

Abbildung 47-17
Das vom Skript erstellte Dialogfeld "Datendatei öffnen"




```
Sub Main
  Call BuildDialog
End Sub

 'Dialogfeld definieren
Sub BuildDialog
  Begin Dialog UserDialog 580,70,"Open Data File",.DialogFunc
    Text 40,7,280,21,"Data file to open:",.txtDialogTitle
    TextBox 40,28,340,21,.txtFilename
    OKButton 470,7,100,21,.cmdOK
    CancelButton 470,35,100,21,.cmdCancel
  End Dialog
  Dim dlg As UserDialog
  Dialog dlg
```

```
End Sub

  'Funktion definieren, die das Verhalten des Dialogfelds festlegt
Function DialogFunc(strDlgItem As String, intAction As Integer, intSuppValue As Integer) As Boolean
   Select Case intAction
   Case 1' Signalton bei der Initialisierung des Dialogfelds
      Beep
   Case 2' Wertänderung oder Klicken auf Schaltfläche
      Select Case strDlgItem
   Case "cmdOK"'Wenn der Benutzer auf "OK" klickt, Datendatei mit angegebenem Dateinamen öffnen
      strFilename = DlgText("txtFilename")
   Call OpenDataFile(strFilename)
   DialogFunc = False
   Case "cmdCancel"'Wenn Benutzer auf "Abbrechen" klickt, Dialogfeld schließen
        DialogFunc = False
      End Select
   End Select
End Function

Sub OpenDataFile(strFilename As Variant)'Datendatei mit angegebenem Dateinamen öffnen
   Dim objDataDoc As ISpssDataDoc
   Set objDataDoc = objSpssApp.OpenDataDoc(strFilename)
End Sub
```

Diese Beispiele sind auch über die Online-Hilfe verfügbar. Sie können diese selbst ausprobieren, indem Sie den Code aus der Hilfe in das Skript-Fenster einfügen.

Fehlersuche in Skripts

Mit dem Menü “Debug” können Sie Ihren Code Schritt für Schritt durchlaufen, wobei Sie jeweils nur eine Zeile oder eine Subroutine ausführen und das Ergebnis betrachten können. Sie können einen Haltepunkt in das Skript einfügen, um die Ausführung bei der Zeile anhalten zu lassen, die den Haltepunkt enthält.

Sie suchen Fehler in einem Autoskript, indem Sie die Autoskript-Datei in einem Skript-Fenster öffnen, Haltepunkte in die Prozedur einfügen, die Sie nach Fehlern absuchen möchten, und dann die statistische Prozedur ausführen lassen, die das Autoskript auslöst.

Einzelschritt. Aktuelle Zeile ausführen. Wenn die aktuelle Zeile eine Subroutine oder ein Funktionsaufruf ist, wird die Ausführung bei der ersten Zeile dieser Subroutine oder Funktion angehalten.

Prozedurschritt. Bis zur nächsten Zeile ausführen. Wenn die aktuelle Zeile eine Subroutine oder ein Funktionsaufruf ist, wird die Subroutine oder Funktion komplett ausgeführt.

Prozedur verlassen. Verlassen der aktuellen Subroutine oder des Funktionsaufrufs.

Bis zur aktuellen Zeile ausführen. Ausführen bis zur aktuellen Zeile.

Haltepunkt ein/aus. Einen Haltepunkt einfügen oder entfernen. Das Skript wird am Haltepunkt angehalten, und das Debug-Fenster wird angezeigt.

Aktuellen Wert anzeigen. Den Wert des aktuellen Ausdrucks anzeigen.

Überwachung hinzufügen. Dem Überwachungsfenster den aktuellen Ausdruck hinzufügen.

Objektkatalog. Den Objektkatalog anzeigen.

Nächste Anweisung festlegen. Die nächste auszuführende Anweisung bestimmen. Es können nur Anweisungen aus der aktuellen Subroutine oder Funktion ausgewählt werden.

Nächste Anweisung anzeigen. Die nächste auszuführende Anweisung anzeigen.

So durchlaufen Sie ein Skript schrittweise

- Wählen Sie aus dem Menü “Debug” eine der Optionen zum schrittweisen Durchlaufen der Befehle aus. Dabei wird eine Subroutine vollständig durchlaufen.

  Die Registerkarten “Immediate”, “Watch”, “Stack” und “Loaded” werden mit der Debug-Symbolleiste im Skript-Fenster angezeigt.

- Verwenden Sie die Symbolleiste oder die Hotkeys, um mit dem Abarbeiten des Skripts fortzufahren.

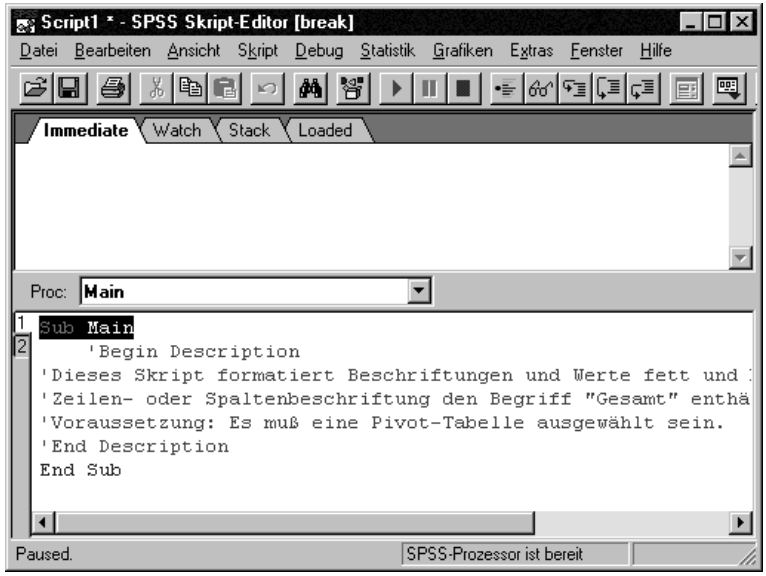
- Wahlweise können Sie Haltepunkt ein/aus auswählen, um einen Haltepunkt an der aktuellen Zeile einzufügen.

  Das Skript wird am Haltepunkt angehalten.

Debug-Fenster (Skripts)

Wenn Sie die Befehlsfolge abarbeiten, werden die Registerkarten “Immediate”, “Watch”, “Stack” und “Loaded” angezeigt.

Abbildung 47-18
Im Skript-Fenster wird das Debug-Fenster angezeigt


Script1 * - SPSS Skript-Editor [break]
Datei Bearbeiten Ansicht Skript Debug Statistik Grafiken Extras Fenster Hilfe
Immediate Watch Stack Loaded
Proc: Main
1 Sub Main
2 'Begin Description
'Dieses Skript formatiert Beschriftungen und Werte fett und
'Zeilen- oder Spaltenbeschriftung den Begriff "Gesamt" enthä
'Voraussetzung: Es muß eine Pivot-Tabelle ausgewählt sein.
'End Description
End Sub
Paused.
SPSS-Prozessor ist bereit

Registerkarte “Immediate”. Klicken Sie auf den Namen einer Variablen und dann auf die Schaltfläche mit der Brille, um den aktuellen Wert der Variablen anzeigen zu lassen. Sie können außerdem einen Ausdruck berechnen, eine Variable zuweisen oder eine Subroutine aufrufen.

- Geben Sie ?ausdruck ein und drücken Sie die Eingabetaste, um den Wert von *ausdruck* anzeigen zu lassen.
- Geben Sie var = ausdruck ein und drücken Sie die Eingabetaste, um den Wert von *var* zu ändern.
- Geben Sie subname args ein und drücken Sie die Eingabetaste, um eine Subroutine oder einen internen Befehl aufzurufen.
- Geben Sie Trace ein und drücken Sie die Eingabetaste, um zum Trace-Modus zu wechseln. Im Trace-Modus wird bei der Abarbeitung eines Skripts jede Anweisung im Direktfenster ausgegeben.

Registerkarte "Watch". Klicken Sie auf diese Registerkarte und wählen Sie Überwachung hinzufügen im Menü "Debug", um eine Variable, eine Funktion oder einen Ausdruck anzeigen zu lassen. Die angezeigten Werte werden jedesmal aktualisiert, wenn die Abarbeitung angehalten wird. Sie können den Ausdruck links von -> bearbeiten. Drücken Sie die Eingabetaste, um alle Werte sofort zu aktualisieren. Drücken Sie Strg+Y, um die Zeile zu löschen.

Registerkarte "Stack". Hiermit werden die Zeilen angezeigt, mit denen die aktuelle Anweisung aufgerufen wurde. Die erste Zeile ist die aktuelle Anweisung, die zweite Zeile ist die, mit der die erste aufgerufen wurde usw. Klicken Sie auf eine beliebige Zeile, um diese im Bearbeitungsfenster zu markieren.

Registerkarte "Loaded". Hier werden die aktuellen aktivierten Skripts aufgelistet. Klicken Sie auf eine Zeile, um sich das Skript anzeigen zu lassen.

Skriptdateien und Syntaxdateien

Syntaxdateien (**.sps*) sind keine Skriptdateien (**.sbs*). Syntaxdateien bestehen aus Befehlen, die in der Befehlssprache geschrieben sind und die Ihnen das Ausführen von statistischen Prozeduren und das Transformieren von Daten ermöglichen. Mit Skripts verfügen Sie über die Möglichkeit, Ausgaben zu beeinflussen und Abläufe zu automatisieren, die Sie sonst mit Hilfe der grafischen Oberfläche der Menüs und Dialogfelder ausführen würden. Mit der Befehlssprache verfügen Sie dagegen über eine Methode, direkt mit dem Ausgabemodul des Programms zu kommunizieren, dem Teil des Systems, in dem die statistischen Berechnungen und die Transformationen von Daten bearbeitet werden.

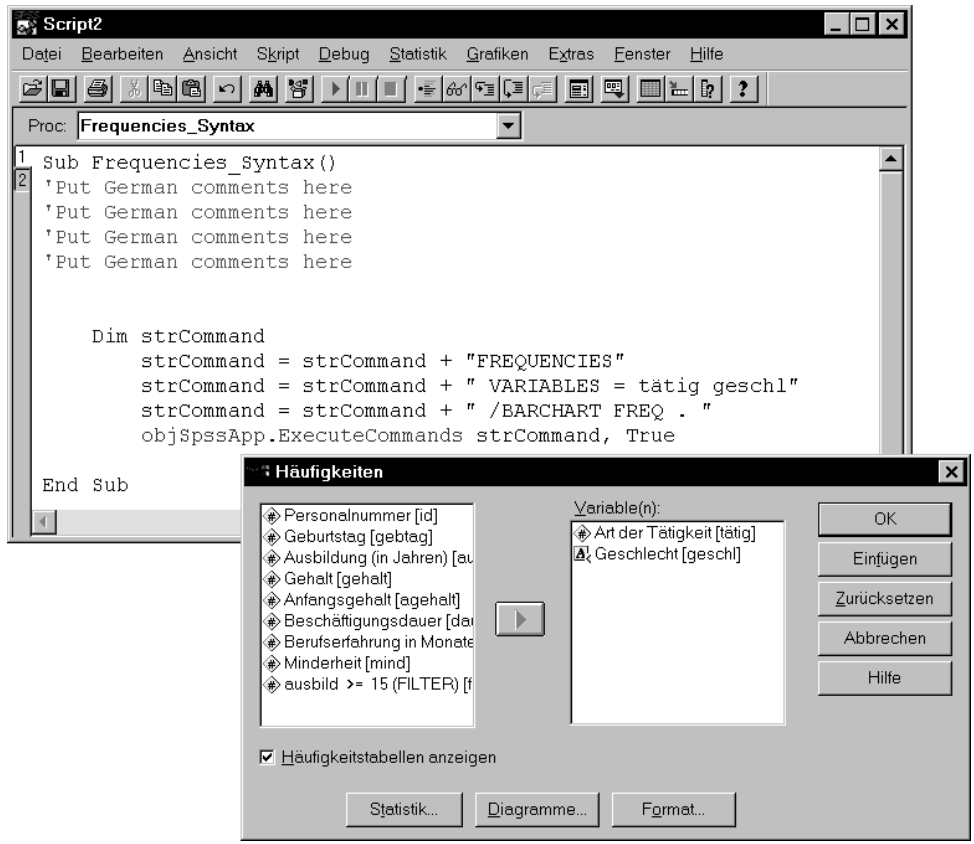
Sie können Skripts und Syntaxdateien kombinieren, um eine noch größere Flexibilität zu erreichen, indem Sie aus der Befehlssyntax heraus ein Skript starten oder Befehlssyntax in einem Skript einbetten.

Ausführen von Befehlssyntax aus einem Skript

Sie können die Befehlssyntax aus einem Skript zur Automatisierung heraus ausführen, indem Sie die Methode `ExecuteCommands` verwenden. Mit der Befehlssyntax verfügen Sie über die Möglichkeit, Datentransformationen durchzuführen, statistische Prozeduren abzuarbeiten und Diagramme zu erstellen. Ein Großteil dieser Funktionen kann nicht direkt aus Befehlsskripts heraus automatisiert werden.

Die einfachste Art, eine Befehlssyntax-Datei zu erstellen, besteht darin, in Dialogfeldern Einstellungen vorzunehmen und die Syntax für diese Angaben in das Skript-Fenster zu kopieren.

Abbildung 47-19
Einfügen von Befehlssyntax in ein Skript



Wenn Sie Dialogfelder unter Verwendung der Menüs des Skript-Fensters öffnen, können Sie mit der Schaltfläche “Einfügen” den gesamten Code einfügen, der zum Ausführen von Befehlen aus Skripts heraus benötigt wird.

Anmerkung: Sie müssen die Menüs des Skript-Fensters verwenden, um das Dialogfeld zu öffnen, sonst werden die Befehle in ein Syntax-Fenster und nicht in das Fenster zum Erstellen eines Skripts eingefügt.

Einfügen von SPSS-Befehlssyntax in ein Skript

- Wählen Sie aus den Menüs “Statistik”, “Grafiken” und “Extras” des Skript-Fensters Befehle zum Öffnen von Dialogfeldern aus.
- Nehmen Sie im Dialogfeld Einstellungen vor.
- Klicken Sie auf Einfügen.

  Anmerkung: Sie müssen die Menüs des Skript-Fensters verwenden, um das Dialogfeld zu öffnen, sonst werden die Befehle in ein Syntax-Fenster und nicht in das Fenster zum Erstellen eines Skripts eingefügt.

Ausführen eines Skripts aus der Befehlssyntax

Sie können den Befehl `SCRIPT` verwenden, um ein Skript aus der Befehlssyntax heraus auszuführen. Legen Sie den Namen des Skripts, das Sie ausführen möchten, wie folgt fest, wobei der Dateiname in Anführungszeichen eingeschlossen wird:

```
SCRIPT 'C:\PROGRAMME\SPSS\CLEAN NAVIGATOR.SBS'.
```

Vollständige Informationen zur Syntax finden Sie im *SPSS Syntax Reference Guide*.

Kapitel
48

Ausgabeverwaltungssystem (OMS)

Das Ausgabeverwaltungssystem (Output Management System, OMS) bietet die Möglichkeit, ausgewählte Ausgabekategorien automatisch in verschiedene Ausgabedateien in unterschiedlichen Formaten schreiben zu lassen. Zu diesen Formaten gehören:

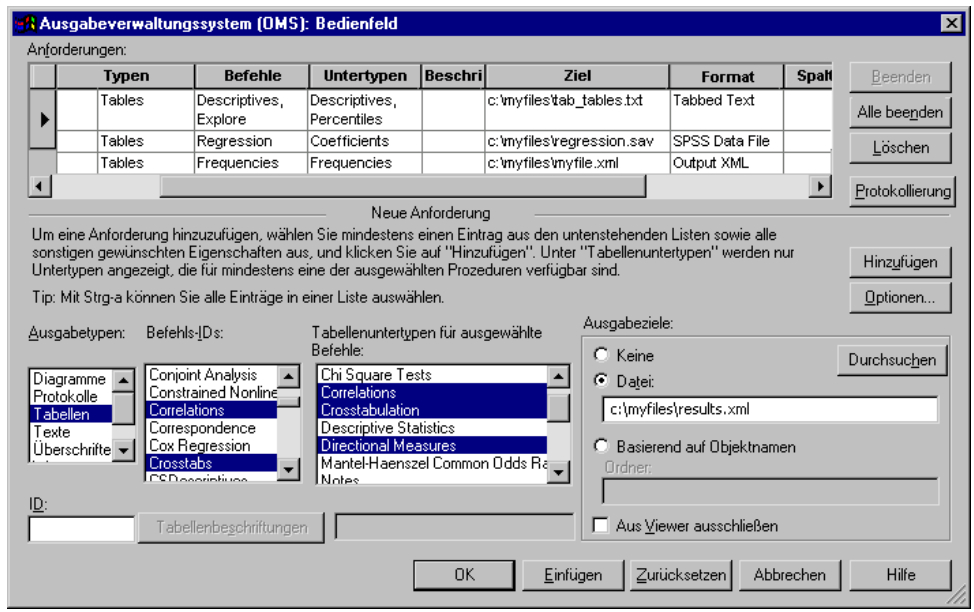
- **SPSS-Datendateiformat (.sav).** Ausgaben, die im Viewer als Pivot-Tabellen angezeigt würden, können in Form einer SPSS-Datendatei geschrieben werden, wodurch die Verwendung der Ausgabe als Eingabe für die darauffolgenden Befehle ermöglicht wird.
- **XML.** Tabellen, Textausgaben und sogar viele Diagramme können im XML-Format geschrieben werden.
- **HTML.** Tabellen und Textausgaben können im HTML-Format geschrieben werden. Standarddiagramme (nicht interaktive Diagramme) und Baummodelldiagramme (Option “Klassifizierungsbaum) können als Bilddateien eingefügt werden. Die Bilddateien werden in einem separaten Unterverzeichnis (Ordner) gespeichert.
- **Text.** Tabellen und Textausgaben können als Tabulator- oder Leerzeichen-getrennter Text geschrieben werden.

So verwenden Sie das Bedienfeld des Ausgabeverwaltungssystems

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Extras
  OMS-Steuerung...

Abbildung 48-1
Systemsteuerung des Ausgabeverwaltungssystems



Mit dem Bedienfeld können Sie die Weiterleitung der Ausgaben an verschiedene Ziele starten und anhalten.

- Alle OMS-Anfragen bleiben aktiv, bis sie ausdrücklich beendet werden oder die Sitzung endet.
- Eine in einer OMS-Anfrage angegebene Zieldatei steht für andere SPSS-Prozeduren und andere Anwendungen so lange nicht zur Verfügung, bis die OMS-Anfrage beendet ist.
- Wenn eine OMS-Anfrage aktiv ist, werden die angegebenen Zieldateien im Arbeitsspeicher (RAM) abgelegt. Aktive OMS-Anfragen, bei denen große Ausgabemengen in externe Dateien geschrieben werden, können somit große Mengen an Arbeitsspeicher belegen.
- Mehrere gleichzeitig ausgeführte OMS-Anfragen sind unabhängig voneinander. Eine Ausgabe kann in unterschiedlichen Formaten an verschiedene Positionen weitergeleitet werden, je nach den Spezifikationen in den einzelnen OMS-Anfragen.

- Die Ausgabeobjekte werden in der Reihenfolge am jeweiligen Ziel ausgegeben, in der sie erstellt wurden. Diese Reihenfolge bei der Erstellung ergibt sich aus der Reihenfolge und der Nutzung der Prozedur, die die Ausgabe erzeugt.
- OMS kann keine Diagramme oder Warnungensammlungen weiterleiten, die durch interaktive Grafikprozeduren erstellt wurden (Menü "Grafiken", Untermenü "Interaktiv"). Dies gilt auch für Karten aus den Kartenprozeduren (Menü "Grafiken", Untermenü "Karten").

So fügen Sie neue OMS-Anfragen hinzu

▶ Wählen Sie die zu berücksichtigenden Ausgabetypen aus (z. B. Tabellen, Diagramme usw.). (Für weitere Informationen siehe "Ausgabeobjekttypen" auf S. 751.)

▶ Wählen Sie die einzuschließenden Befehle aus. Soll die gesamte Ausgabe berücksichtigt werden, wählen Sie alle Elemente in der Liste aus. (Für weitere Informationen siehe "Befehls-IDs und Tabellenuntertypen" auf S. 753.)

▶ Bei Befehlen, die eine Pivot-Tabelle als Ausgabe erzeugen, wählen Sie die zugehörigen Tabellentypen aus.

Die Liste enhält nur die Tabellen, die in den ausgewählten Befehlen zur Verfügung stehen. Es werden sämtliche Tabellentypen aufgeführt, die in mindestens einem der ausgewählten Befehle in der Liste verfügbar sind. Falls Sie keine Befehle ausgewählt haben, werden alle Tabellentypen aufgeführt. (Für weitere Informationen siehe "Befehls-IDs und Tabellenuntertypen" auf S. 753.)

▶ Wenn Sie Tabellen nicht nach den Untertypen, sondern nach dem Beschriftungstext auswählen möchten, klicken Sie auf Beschriftungen. (Für weitere Informationen siehe "Labels" auf S. 755.)

▶ Klicken Sie auf Optionen, und bestimmen Sie das Ausgabeformat (z. B. SPSS-Datendatei, XML, HTML). (Standardmäßig wird bei der Ausgabe das XML-Format verwendet. Für weitere Informationen siehe "OMS: Options" auf S. 757.)

▶ Geben Sie ein Ausgabeziel an:

- **Datei.** Alle ausgewählten Ausgaben werden in eine Datei weitergeleitet.
- **Basierend auf Objektnamen.** Die Ausgabe wird anhand der Objektnamen an mehrere Zieldateien weitergeleitet. Für jedes Ausgabeobjekt wird eine separate Datei erstellt. Die Dateinamen basieren entweder auf den Namen von

Tabellenuntertypen oder den Beschriftungen der Tabellen. Geben Sie den Namen des Zielordners ein.

- **Neues Daten-Set.** Bei Ausgaben im Format von SPSS-Datendateien können Sie die Ausgabe in ein Daten-Set weiterleiten. Das Daten-Set ist für die anschließende Verwendung in derselben Sitzung verfügbar. Es wird jedoch nicht gespeichert, sofern Sie es nicht ausdrücklich vor dem Beenden der Sitzung als Datei speichern. Diese Option ist nur für Ausgaben im Format von SPSS-Datendateien verfügbar. Die Namen von Daten-Sets müssen den SPSS-Regeln zum Benennen von Variablen entsprechen. Für weitere Informationen siehe "Variablennamen" in Kapitel 5 auf S. 99.

▶ Die folgenden Optionen sind verfügbar:

- Schließen Sie die ausgewählten Ausgaben aus dem Viewer aus. Mit Aus Viewer ausschließen werden die Ausgabetypen in der OMS-Anfrage nicht im Viewer-Fenster angezeigt. Enthalten mehrere aktive OMS-Anfragen dieselben Ausgabetypen, wird anhand der jüngsten OMS-Anfrage mit den betreffenden Ausgabetypen entschieden, ob die Ausgabetypen angezeigt werden sollen oder nicht. Für weitere Informationen siehe "Ausschließen der Ausgabeanzeige aus dem Viewer" auf S. 763.
- Weisen Sie der Anfrage einen ID-String zu. Alle Anfragen erhalten automatisch einen ID-Wert, und Sie können den Standard-ID-String durch eine aussagekräftige ID ersetzen. Dies ist insbesondere dann von Nutzen, wenn mehrere aktive Anfragen vorliegen, die Sie einfach unterscheiden möchten. Die zugewiesenen ID-Werte dürfen nicht mit einem Dollar-Zeichen ($) beginnen.

Hier finden Sie einige Tipps, wie Sie mehrere Elemente in einer Liste auswählen:

- Drücken Sie Strg+A, um alle Elemente in einer Liste gleichzeitig auszuwählen.
- Wenn Sie bei gedrückter Umschalttaste klicken, können Sie mehrere aufeinander folgende Elemente auswählen.
- Wenn Sie bei gedrückter Strg-Taste klicken, können Sie mehrere nicht aufeinander folgende Elemente auswählen.

So beenden und löschen Sie OMS-Anfragen

Aktive und neue OMS-Anfragen werden in der Liste "Anforderungen" aufgeführt. Die neueste Anfrage befindet sich dabei an oberster Stelle. Die Breite der Datenspalten kann geändert werden. Klicken Sie hierzu auf die Spaltenbegrenzungen,

und ziehen Sie sie an die gewünschte Position. Darüber hinaus können Sie in horizontaler Richtung durch die Liste blättern und so weitere Informationen zu einer bestimmten Anfrage anzeigen lassen.

Ein Sternchen (*) neben dem Wort Aktiv in der Spalte *Status* bedeutet, dass die zugehörige OMS-Anfrage mit Befehlssyntax erzeugt wurde, die im Bedienfeld nicht zur Verfügung steht.

So beenden Sie eine bestimmte aktive OMS-Anfrage:

- Klicken Sie in der Liste "Anforderungen" auf eine beliebige Zelle in der Zeile dieser Anfrage.
- Klicken Sie auf Beenden.

So beenden Sie alle aktiven OMS-Anfragen:

- Klicken Sie auf Alle beenden.

So löschen Sie eine neue Anfrage (eine hinzugefügte Anfrage, die noch nicht aktiv ist):

- Klicken Sie in der Liste "Anforderungen" auf eine beliebige Zelle in der Zeile dieser Anfrage.
- Klicken Sie auf Löschen.

Anmerkung: Aktive OMS-Anfragen werden erst dann beendet, wenn Sie auf OK klicken.

Ausgabeobjekttypen

Es gibt sieben verschiedene Arten von Ausgabeobjekten:

Diagramme. Diagramme (mit Ausnahme von "interaktiven" Diagrammen und Karten). Diagrammobjekte werden nur bei den Zielformaten XML und HTML berücksichtigt. Beim HTML-Format werden die Bilddateien in einem separaten Unterverzeichnis (Ordner) gespeichert.

Protokolle. Protokolltextobjekte. Protokollobjekte enthalten bestimmte Arten von Fehler- und Warnmeldungen. Je nach den Einstellungen unter "Optionen" (Menü "Bearbeiten", "Optionen", Registerkarte "Viewer") umfassen die Log-Objekte

möglicherweise auch die Befehlssyntax, die während der Sitzung ausgeführt wurde. Protokollobjekte werden im Gliederungsfenster des Viewers mit *Log* gekennzeichnet.

Tabellen. Ausgabeobjekte, die im Viewer als Pivot-Tabellen dargestellt werden (einschließlich der Anmerkungstabellen). Tabellen sind die einzigen Ausgabeobjekte, die im SPSS-Datendateiformat (*.sav*) weitergeleitet werden können.

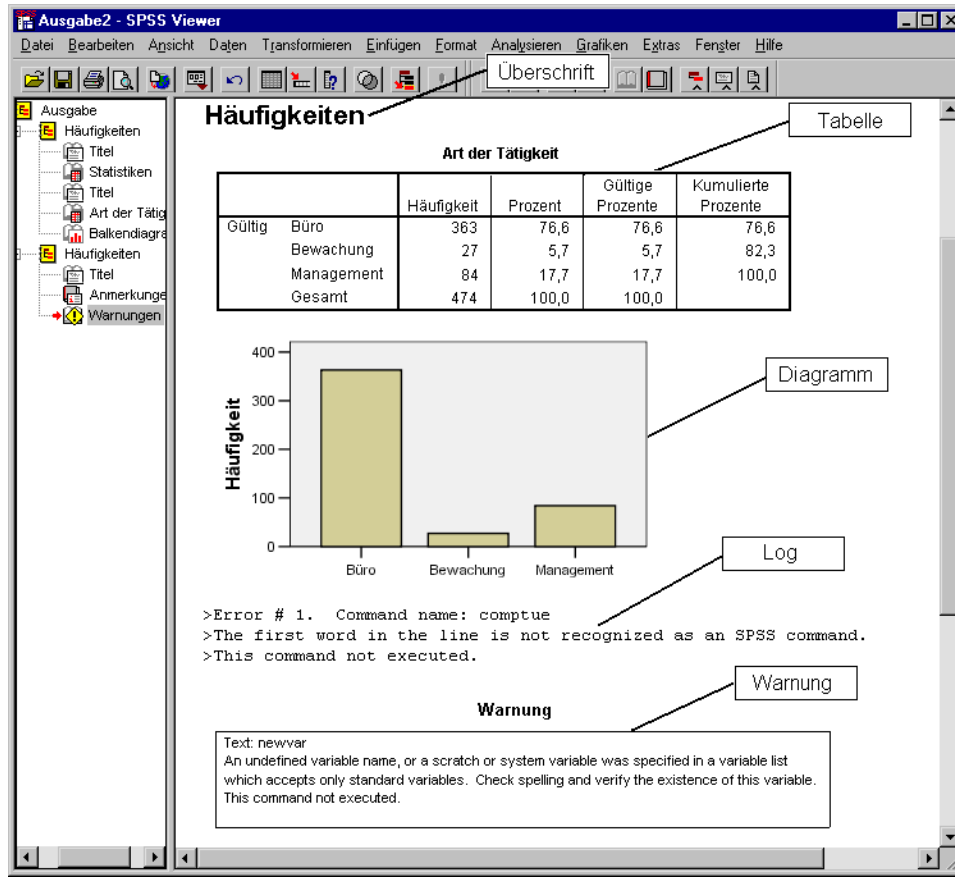
Text. Textobjekte, die weder Logs noch Überschriften sind (einschließlich der Objekte, die im Gliederungsfenster des Viewers mit *Textausgabe* gekennzeichnet sind).

Bäume. Baummodelldiagramme, die mit der Option "Classification Tree" erzeugt wurden. Baumobjekte werden nur bei den Zielformaten XML und HTML berücksichtigt.

Überschriften. Textobjekte, die im Gliederungsfenster des Viewers mit *Titel* gekennzeichnet sind. Beim Output XML-Format werden Überschriften-Textobjekte nicht berücksichtigt.

Warnungen. Warnungensammlungen. Warnungensammlungen enthalten bestimmte Arten von Fehler- und Warnmeldungen.

Abbildung 48-2
Ausgabeobjekttypen



Befehls-IDs und Tabellenuntertypen

Befehls-IDs

Befehls-IDs sind für alle Statistik- und Diagrammprozeduren verfügbar, außerdem für alle Befehle, bei denen Ausgabeblöcke mit eigener identifizierbarer Überschrift im Gliederungsfenster des Viewers erzeugt werden. Diese IDs sind in der Regel (jedoch nicht immer) identisch oder nahezu identisch mit den Namen der Prozeduren in den

Menüs und den Dialogfeldtiteln, die wiederum in der Regel (jedoch nicht immer) identisch oder nahezu identisch mit den Bezeichnungen der zugrunde liegenden SPSS-Befehlsnamen sind. Die Befehls-ID für die Prozedur "Häufigkeiten" lautet beispielsweise "Häufigkeiten"; dies ist auch die Bezeichnung des zugrunde liegenden Befehls.

In einigen Fällen weisen der Name der Prozedur und die Befehls-ID und/oder der Befehlsname allerdings beträchtliche Unterschiede auf. Beispielsweise greifen alle Prozeduren im Untermenü "Nichtparametrisch" des Menüs "Analysieren" auf denselben zugrunde liegenden Befehl zurück, und die Befehls-ID ist mit dem Namen des zugrundeliegenden Befehls identisch: Npar Tests.

Tabellenuntertypen

Tabellenuntertypen sind die verschiedenen Typen von Pivot-Tabellen, die erstellt werden können. Einige Untertypen werden nur von einem einzigen Befehl erzeugt, andere Untertypen dagegen von mehreren Befehlen (die Tabellen zeigen jedoch unter Umständen nicht dasselbe Erscheinungsbild). Die Namen der Tabellenuntertypen sind normalerweise aussagekräftig. Es können allerdings zahlreiche Untertypen zur Auswahl stehen (insbesondere wenn Sie viele Befehle ausgewählt haben). Zwei Untertypen können auch sehr ähnliche Namen besitzen.

So suchen Sie Befehls-IDs und Tabellenuntertypen

Im Zweifelsfall können Sie die Befehls-IDs und die Namen der Tabellenuntertypen im Viewer-Fenster suchen:

- Starten Sie die Prozedur, um Ausgaben im Viewer zu erzeugen.
- Klicken Sie mit der rechten Maustaste auf das Element im Gliederungsfenster des Viewers.
- Wählen Sie die Option OMS-Befehls-ID kopieren oder die Option OMS-Tabellenuntertyp kopieren.
- Fügen Sie die kopierte Befehls-ID bzw. den Namen des Untertabellentyps in einen Text-Editor ein (z. B. in ein SPSS-Syntax-Fenster).

Labels

Als Alternative zu Namen von Tabellenuntertypen können Sie Tabellen auf der Grundlage des Textes auswählen, der im Gliederungsfenster des Viewers angezeigt wird. Sie können auch andere Objekttypen anhand der Beschriftung auswählen. Beschriftungen helfen beim Unterscheiden zwischen mehreren Tabellen desselben Typs, bei denen der Gliederungstext ein Attribut des jeweiligen Ausgabeobjekts angibt, beispielsweise die Variablennamen oder -beschriftungen. Es gibt jedoch eine Reihe von Faktoren, die den Labeltext beeinflussen können:

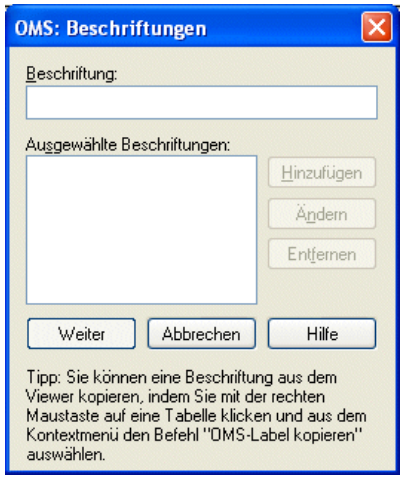
- Wenn die Verarbeitung aufgeteilter Dateien aktiviert ist, kann die Gruppen-ID für die aufgeteilte Datei an die Beschriftung angehängt werden.
- Beschriftungen mit Informationen zu Variablen oder Werten sind abhängig von den aktuellen Einstellungen für die Beschriftung der Ausgabe (Menü “Bearbeiten”, “Optionen”, Registerkarte “Beschriftung der Ausgabe”).
- Labels richten sich außerdem nach der aktuellen Einstellung für die Ausgabesprache (Menü “Bearbeiten”, “Optionen”, Registerkarte “Allgemein”).

So legen Sie Beschriftungen zum Identifizieren von Ausgabetabellen fest

▶ Wählen Sie im Bedienfeld des Ausgabeverwaltungssystems mindestens einen Ausgabetyp und anschließend mindestens einen Befehl aus.

▶ Klicken Sie auf Variablenlabel.

Abbildung 48-3
Dialogfeld "OMS: Beschriftungen"



- Geben Sie die Beschriftung *auf dieselbe Weise* ein, wie sie im Gliederungsfenster des Viewer-Fensters aufgeführt wird. (Alternativ können Sie mit der rechten Maustaste auf das Element in der Gliederung klicken, die Option OMS-Label kopieren auswählen und die kopierte Beschriftung dann im Textfeld "Beschriftung" einfügen.)
- Klicken Sie auf Hinzufügen.
- Wiederholen Sie diesen Vorgang für jede Beschriftung, die Sie hinzufügen möchten.
- Klicken Sie auf Weiter.

Platzhalter

Sie können ein Sternchen (*) als Platzhalterzeichen als letztes Zeichen im Bezeichnungs-String verwenden. Alle Beschriftungen, die mit dem angegebenen String beginnen (alle Zeichen mit Ausnahme des Sternchens), werden ausgewählt. Dies ist nur dann möglich, wenn das Sternchen das letzte Zeichen ist, weil Sternchen durchaus als zulässige Zeichen innerhalb einer Beschriftung auftreten können.

OMS: Options

Das Dialogfeld “OMS: Optionen” bietet die folgenden Möglichkeiten:

- Legen Sie das Ausgabeformat fest.
- Schließen Sie die Diagramm- und Baummodelldiagrammausgabe ein oder aus, und bestimmen Sie das Grafikformat.
- Geben Sie an, welche Tabellendimensionselemente in die Zeilendimension eingehen sollen.
- Schließen Sie beim SPSS-Datendateiformat eine Variable für die laufende Tabellennummer ein, die als Quelle für die einzelnen Fälle herangezogen werden soll.

So legen Sie OMS-Optionen fest

▶ Klicken Sie auf Optionen in der Systemsteuerung des Ausgabeverwaltungssystems (OMS).

Abbildung 48-4
Dialogfeld “OMS: Optionen”

Format

Ausgabe-XML. XML, das dem Schema *spss-output* entspricht. Standarddiagramme werden als XML berücksichtigt, das dem Schema *vizml* entspricht. Darüber hinaus können Sie alle Diagramme und Karten als separate Dateien im ausgewählten Grafikformat exportieren.

HTML. Ausgabeobjekte, die als Pivot-Tabellen im Viewer dargestellt würden, werden in einfache HTML-Tabellen umgewandelt. Attribute aus Tabellenvorlagen (z. B. Schriftart, Rahmenformate, Farben) werden nicht unterstützt. Textausgabeobjekte werden als <PRE> in der HTML gekennzeichnet. Wenn Sie die Diagramme berücksichtigen lassen, werden diese als separate Dateien im ausgewählten Grafikformat exportiert und als Verweis (<IMG SRC = 'Dateiname.ext'>) in das HTML-Dokument eingebettet. Die Bilddateien werden in einem separaten Unterverzeichnis (Ordner) gespeichert.

SPSS-Datendatei. Hierbei handelt es sich um ein binäres Dateiformat. Alle Ausgabeobjekttypen mit Ausnahme der Tabellen sind ausgeschlossen. Jede Spalte einer Tabelle wird zu einer Variablen in der Datendatei. Soll eine mit OMS erstellte Datendatei noch in derselben Sitzung genutzt werden, beenden Sie die aktive OMS-Anfrage, damit Sie die Datendatei öffnen können. Für weitere Informationen siehe “Weiterleiten der Ausgabe an SPSS-Datendateien” auf S. 763.

SVWS XML. XML, das von SmartViewer-Web-Server verwendet wird. Dies ist eigentlich eine JAR/ZIP-Datei mit XML-, CSV- und anderen Dateien. Der SmartViewer-Web-Server ist ein separat erhältliches, servergestütztes Produkt.

Text. Text, der mit Leerzeichen getrennt ist. Die Ausgabe wird als Text geschrieben. Bei Schriftarten mit fester Breite wird die Tabellenausgabe mit Hilfe von Leerzeichen ausgerichtet. Alle Diagramme und Karten werden ausgeschlossen.

Tabulatorgetrennter Text. Text, der mit Tabulatoren getrennt ist. Bei Ausgaben, die im Viewer als Pivot-Tabellen angezeigt werden, begrenzen die Tabulatoren die Tabellenspaltenelemente. Textblockzeilen werden unverändert geschrieben; der Text wird nicht mit Tabulatoren an sinnvollen Positionen gegliedert. Alle Diagramme und Karten werden ausgeschlossen.

Grafiken

Beim HTML-Format können Sie Diagramme (mit Ausnahme von interaktiven Diagrammen) und Baummodelldiagramme als Bilddateien einschließen. Für jedes Diagramm und/oder für jeden Baum wird eine separate Bilddatei erstellt, und für die einzelnen Bilddateien werden jeweils die Standard-Tags <IMG SRC='Dateiname'> in das HTML aufgenommen.

Die Bilddateien werden in einem separaten Unterverzeichnis (Ordner) gespeichert. Der Name des Unterverzeichnisses ist der Name der HTML-Zieldatei, ohne Dateinamenerweiterung, dafür mit dem Suffix *_files*. Wenn die HTML-Zieldatei beispielsweise den Namen *julidaten.htm* besitzt, erhält das Bildunterverzeichnis den Namen *julidaten_files*.

- **Format.** Als Bildformate stehen PNG, JPEG, EMF und BMP zur Verfügung.
- **Größe.** Sie können das Bild von 10 % bis 200 % skalieren.

Tabellen-Pivots

Bei der Ausgabe von Pivot-Tabellen können Sie das oder die Dimensionselemente bestimmen, die in den Spalten auftreten sollen. Alle anderen Dimensionselemente treten in den Zeilen auf. Beim SPSS-Datendateiformat werden die Tabellenspalten zu Variablen und die Zeilen zu Fällen.

- Wenn Sie mehrere Dimensionselemente für die Spalten angeben, werden diese Elemente in der Reihenfolge in den Spalten verschachtelt, in der sie aufgeführt sind. Beim SPSS-Datendateiformat werden die Variablennamen aus verschachtelten Spaltenelementen gebildet. Für weitere Informationen siehe “Variablennamen in Datendateien aus dem OMS” auf S. 772.
- Wenn eine Tabelle keine der aufgeführten Dimensionselemente enthält, werden alle Dimensionselemente dieser Tabelle in den Zeilen aufgeführt.
- Die hier angegebenen Tabellen-Pivots wirken sich nicht auf die Tabellen aus, die im Viewer dargestellt werden.

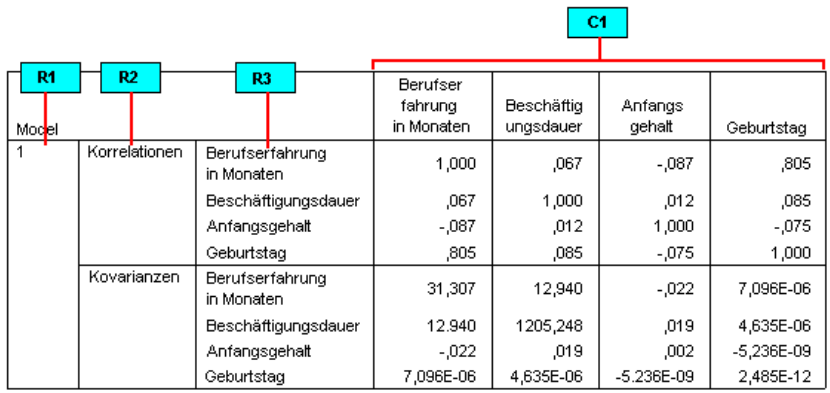
Jede Dimension einer Tabelle (Zeile, Spalte, Schicht) kann null oder mehr Elemente enthalten. Eine einfache Kreuztabelle mit zwei Dimensionen enthält beispielsweise ein einziges Zeilendimensionselement und ein einziges Spaltendimensionselement, die jeweils eine der in der Tabelle verwendeten Variablen enthalten. Die Dimensionselemente für die Spaltendimension können wahlweise mit Hilfe von Positionsargumenten oder mit den Dimensionselement-”Namen” festgelegt werden.

Alle Dimensionen in einer einzelnen Zeile. Hiermit wird eine einzelne Zeile für jede Tabelle erstellt. Bei Datendateien im SPSS-Format bedeutet dies, dass jede Tabelle einen einzelnen Fall darstellt und alle Tabellenelemente Variablen sind.

Liste der Positionen. Ein Positionsargument besteht in der Regel aus einem Buchstaben für die Standardposition des Elements (C für Spalte, R für Zeile, L für Schicht), gefolgt von einer positiven Ganzzahl, aus der die Standardposition innerhalb dieser Dimension hervorgeht. R1 bezeichnet beispielsweise das äußerste Zeilendimensionselement.

- Sollen mehrere Elemente aus mehreren Dimensionen angegeben werden, trennen Sie die einzelnen Dimensionen jeweils mit einem Leerzeichen, z. B. R1 C2.
- Steht nach dem Dimensionsbuchstaben die Zeichenfolge "ALL", bedeutet dies, daß alle Elemente in der betreffenden Dimension in ihrer Standardreihenfolge berücksichtigt werden. CALL entspricht beispielsweise dem Standardverhalten; die Spalten werden hierbei auf der Grundlage aller Spaltenelemente in ihrer Standardreihenfolge gebildet.
- Mit CALL RALL LALL (oder RALL CALL LALL usw.) werden alle Dimensionselemente in die Spalten aufgenommen. Beim SPSS-Datendateiformat entsteht hierbei je eine Zeile/ein Fall pro Tabelle in der Datendatei.

Abbildung 48-5
Zeilen- und Spaltenpositionsargumente



Modell			Berufserfahrung in Monaten	Beschäftigungsdauer	Anfangsgehalt	Geburtstag
1	Korrelationen	Berufserfahrung in Monaten	1,000	,067	-,087	,805
		Beschäftigungsdauer	,067	1,000	,012	,085
		Anfangsgehalt	-,087	,012	1,000	-,075
		Geburtstag	,805	,085	-,075	1,000
	Kovarianzen	Berufserfahrung in Monaten	31,307	12,940	-,022	7,096E-06
		Beschäftigungsdauer	12.940	1205,248	,019	4,635E-06
		Anfangsgehalt	-,022	,019	,002	-5,236E-09
		Geburtstag	7,096E-06	4,635E-06	-5.236E-09	2,485E-12

Liste der Dimensionsnamen. Als Alternative zu Positionsargumenten können Sie die "Namen" der Dimensionselemente verwenden, also die Textbeschriftungen, die in der Tabelle aufgeführt werden. Eine einfache zweidimensionale Kreuztabelle

enthält beispielsweise ein einziges Zeilendimensionselement und ein einziges Spaltendimensionselement, die jeweils mit einer Beschriftung auf der Grundlage der Variablen in diesen Dimensionen versehen sind, außerdem ein einziges Schichtdimensionselement mit der Beschriftung *Statistik* (wenn Deutsch die Ausgabesprache ist).

- Die Dimensionselementnamen sind abhängig von der Ausgabesprache und/oder von den Einstellungen, die sich auf die Anzeige von Variablennamen und/oder Beschriftungen in Tabellen auswirken.
- Jeder Dimensionselementname muss in einfache oder doppelte Anführungszeichen eingeschlossen werden. Sollen mehrere Dimensionselementnamen angegeben werden, trennen Sie die einzelnen, in Anführungsstrichen stehenden Namen jeweils mit einem Leerzeichen.

Die Beschriftungen für die Dimensionselemente sind nicht in jedem Fall deutlich.

So lassen Sie alle Dimensionselemente und deren Beschriftungen für eine Pivot-Tabelle anzeigen

▶ Aktivieren Sie die Tabelle im Viewer durch Doppelklicken.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Ansicht
  Alles einblenden

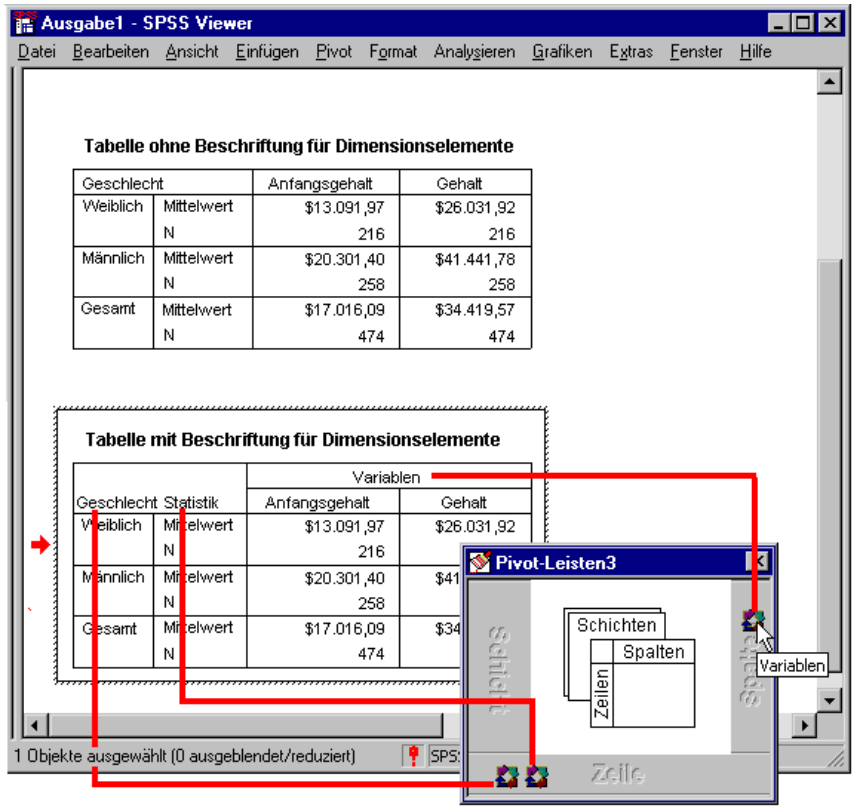
und/oder

▶ Falls die Pivot-Leisten nicht angezeigt werden, wählen Sie folgende Befehle aus den Menüs aus:

Pivot
  Pivot-Leisten

▶ Zeigen Sie mit dem Mauszeiger auf ein Symbol, um die jeweilige Beschriftung anzuzeigen.

Abbildung 48-6
Dimensionselementnamen in der Tabelle und in den Pivot-Leisten



Protokollierung

Sie können die OMS-Aktivitäten in einem Protokoll im XML- oder Textformat protokollieren lassen.

- Im Protokoll werden alle neuen OMS-Anfragen für die aktuelle Sitzung aufgezeichnet, nicht jedoch OMS-Anfragen, die bereits aktiv waren, bevor Sie die Protokollierung aktiviert haben.
- Die aktuelle Protokolldatei wird beendet, sobald Sie eine neue Protokolldatei angeben oder die Option OMS-Aktivität protokollieren deaktivieren.

So aktivieren Sie die OMS-Protokollierung:

- Klicken Sie auf Protokollierung in der Systemsteuerung des Ausgabeverwaltungssystems (OMS).

Ausschließen der Ausgabeanzeige aus dem Viewer

Das Kontrollkästchen Aus Viewer ausschließen unterdrückt die Anzeige aller Ausgaben im Viewer-Fenster, die in der OMS-Anfrage ausgewählt wurden. Dies eignet sich insbesondere für Produktionsjobs, bei denen eine umfangreiche Ausgabe entsteht, ohne dass die Ergebnisse in einem Viewer-Dokument (*.spo*-Datei) dargestellt werden müssen. Darüber hinaus können Sie mit dieser Funktion die Anzeige bestimmter Ausgabeobjekte unterdrücken, die einfach nicht dargestellt werden sollen, ohne andere Ausgaben an eine externe Datei und in einem anderen Format weiterzuleiten.

So unterdrücken Sie die Anzeige bestimmter Ausgabeobjekte, ohne andere Ausgaben an eine externe Datei weiterzuleiten:

- Erstellen Sie eine OMS-Anfrage, mit der die unerwünschte Ausgabe ermittelt wird.
- Wählen Sie Aus Viewer ausschließen.
- Wählen Sie für das Ausgabeziel Datei, aber machen Sie keine Dateiangabe.
- Klicken Sie auf Hinzufügen.

Die ausgewählte Ausgabe wird aus dem Viewer ausgeschlossen; die restliche Ausgabe wird wie gewohnt im Viewer dargestellt.

Weiterleiten der Ausgabe an SPSS-Datendateien

Eine SPSS-Datendatei besteht aus Variablen (in den Spalten) und Fällen (in den Zeilen). Im Wesentlichen entspricht dies dem Verfahren, wie Pivot-Tabellen in Datendateien umgewandelt werden:

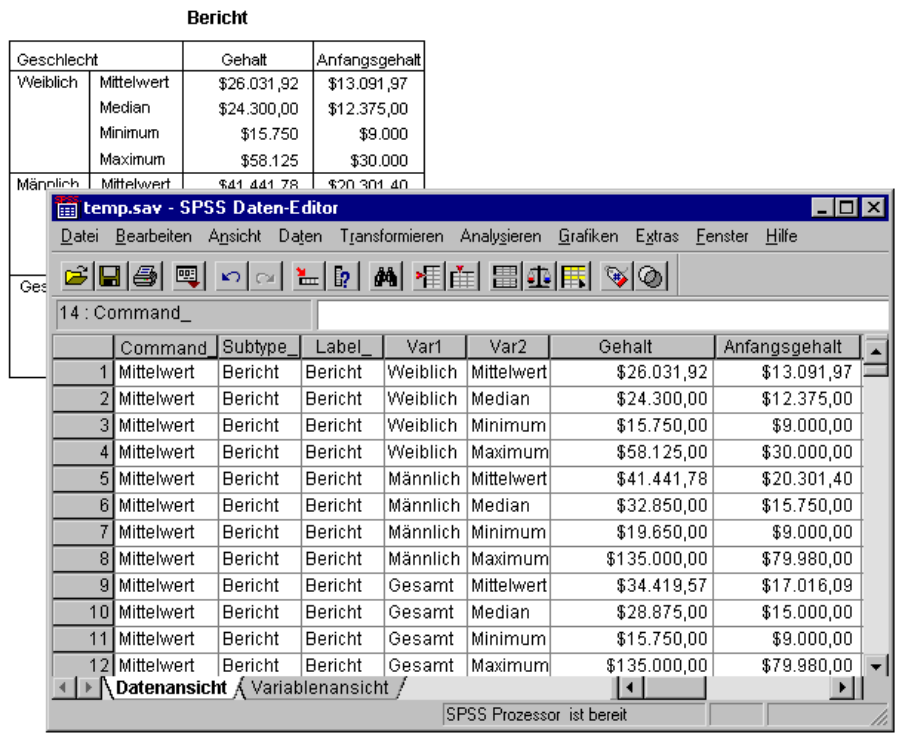
- Spalten in der Tabelle werden zu Variablen in der Datendatei. Aus den Spaltenbeschriftungen werden gültige Variablennamen gebildet.
- Die Zeilenbeschriftungen in der Spalte werden zu Variablen mit generischen Variablennamen (*Var1*, *Var2*, *Var3*) in der Datendatei. Die Werte dieser Variablen entsprechen den Zeilenbeschriftungen in der Tabelle.

- In die Datendatei werden automatisch drei Tabellen-ID-Variablen aufgenommen: *Command_*, *Subtype_* und *Label_*. Alle drei Variablen sind String-Variablen. Die ersten beiden Variablen bezeichnen die Befehls- und die Untertyp-ID. Für weitere Informationen siehe “Befehls-IDs und Tabellenuntertypen” auf S. 753. *Label_* enthält den Tabellentiteltext.
- Zeilen in der Tabelle werden zu Fällen in der Datendatei.

Beispiel: Einzelne zweidimensionale Tabelle

Im einfachsten Fall (also bei einer einzelnen, zweidimensionalen Tabelle) werden die Tabellenspalten zu Variablen und die Zeilen zu Fällen in der Datendatei.

Abbildung 48-7
Einzelne zweidimensionale Tabelle



Bericht

Geschlecht		Gehalt	Anfangsgehalt
Weiblich	Mittelwert	$26.031,92	$13.091,97
	Median	$24.300,00	$12.375,00
	Minimum	$15.750	$9.000
	Maximum	$58.125	$30.000
Männlich	Mittelwert	$41.441,78	$20.301,40

temp.sav - SPSS Daten-Editor

Datei Bearbeiten Ansicht Daten Transformieren Analysieren Grafiken Extras Fenster Hilfe

14 : Command_

	Command_	Subtype_	Label_	Var1	Var2	Gehalt	Anfangsgehalt
1	Mittelwert	Bericht	Bericht	Weiblich	Mittelwert	$26.031,92	$13.091,97
2	Mittelwert	Bericht	Bericht	Weiblich	Median	$24.300,00	$12.375,00
3	Mittelwert	Bericht	Bericht	Weiblich	Minimum	$15.750,00	$9.000,00
4	Mittelwert	Bericht	Bericht	Weiblich	Maximum	$58.125,00	$30.000,00
5	Mittelwert	Bericht	Bericht	Männlich	Mittelwert	$41.441,78	$20.301,40
6	Mittelwert	Bericht	Bericht	Männlich	Median	$32.850,00	$15.750,00
7	Mittelwert	Bericht	Bericht	Männlich	Minimum	$19.650,00	$9.000,00
8	Mittelwert	Bericht	Bericht	Männlich	Maximum	$135.000,00	$79.980,00
9	Mittelwert	Bericht	Bericht	Gesamt	Mittelwert	$34.419,57	$17.016,09
10	Mittelwert	Bericht	Bericht	Gesamt	Median	$28.875,00	$15.000,00
11	Mittelwert	Bericht	Bericht	Gesamt	Minimum	$15.750,00	$9.000,00
12	Mittelwert	Bericht	Bericht	Gesamt	Maximum	$135.000,00	$79.980,00

Datenansicht | Variablenansicht

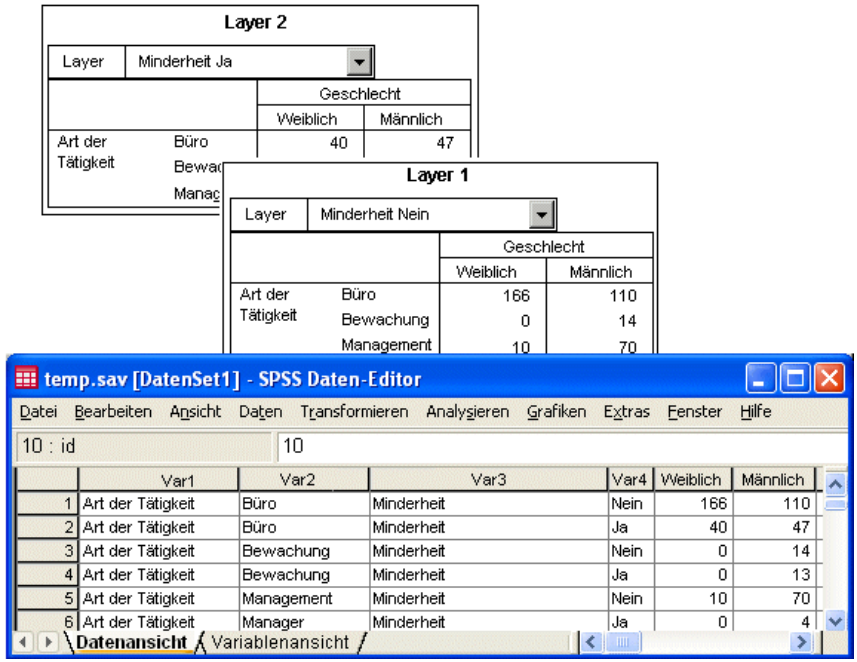
SPSS Prozessor ist bereit

- Die ersten drei Variablen kennzeichnen die Quelltabelle anhand des Befehls, des Untertyps und der Beschriftung.
- Die beiden Elemente, die die Zeilen in der Tabelle definiert hatten (Werte für die Variable *Geschlecht* und statistische Maße) werden mit den generischen Variablennamen *Var1* und *Var2* versehen. Beide Variablen sind String-Variablen.
- Aus den Spaltenbeschriftungen in der Tabelle werden gültige Variablennamen gebildet. In diesem Fall beruhen diese Variablennamen auf den Variablenlabels der drei in der Tabelle ausgewerteten metrischen Variablen. Falls für die Variablen keine Variablenlabels definiert sind oder die Variablennamen anstelle der Variablenlabels als Spaltenbeschriftungen in der Tabelle angezeigt werden sollen, wären die Variablennamen in der neuen Datendatei mit den Namen in der Quelldatendatei identisch.

Beispiel: Tabellen mit Schichten

Neben Zeilen und Spalten kann eine Tabelle eine dritte Dimension aufweisen: die Schichtdimension.

Abbildung 48-8
Tabelle mit Schichten



Layer 2

Layer	Minderheit Ja		
		Geschlecht	
		Weiblich	Männlich
Art der Tätigkeit	Büro	40	47
	Bewac		
	Manag		

Layer 1

Layer	Minderheit Nein		
		Geschlecht	
		Weiblich	Männlich
Art der Tätigkeit	Büro	166	110
	Bewachung	0	14
	Management	10	70

temp.sav [DatenSet1] - SPSS Daten-Editor

Datei Bearbeiten Ansicht Daten Transformieren Analysieren Grafiken Extras Fenster Hilfe

10 : id 10

	Var1	Var2	Var3	Var4	Weiblich	Männlich
1	Art der Tätigkeit	Büro	Minderheit	Nein	166	110
2	Art der Tätigkeit	Büro	Minderheit	Ja	40	47
3	Art der Tätigkeit	Bewachung	Minderheit	Nein	0	14
4	Art der Tätigkeit	Bewachung	Minderheit	Ja	0	13
5	Art der Tätigkeit	Management	Minderheit	Nein	10	70
6	Art der Tätigkeit	Manager	Minderheit	Ja	0	4

Datenansicht Variablenansicht

- In der Tabelle werden die Schichten durch die Variable *Minderheit* definiert. In der Datendatei werden so zwei zusätzliche Variablen erstellt: eine Variable, die das Schichtelement identifiziert, und eine Variable, die die Kategorien des Schichtelements bezeichnet.
- Die Variablen, die aus den Schichtelementen gebildet wurden, sind ebenfalls String-Variablen mit generischen Variablennamen (Präfix *Var*, gefolgt von einer laufenden Nummer), so wie die Variablen aus den Zeilenelementen.

Datendateien aus mehreren Tabellen

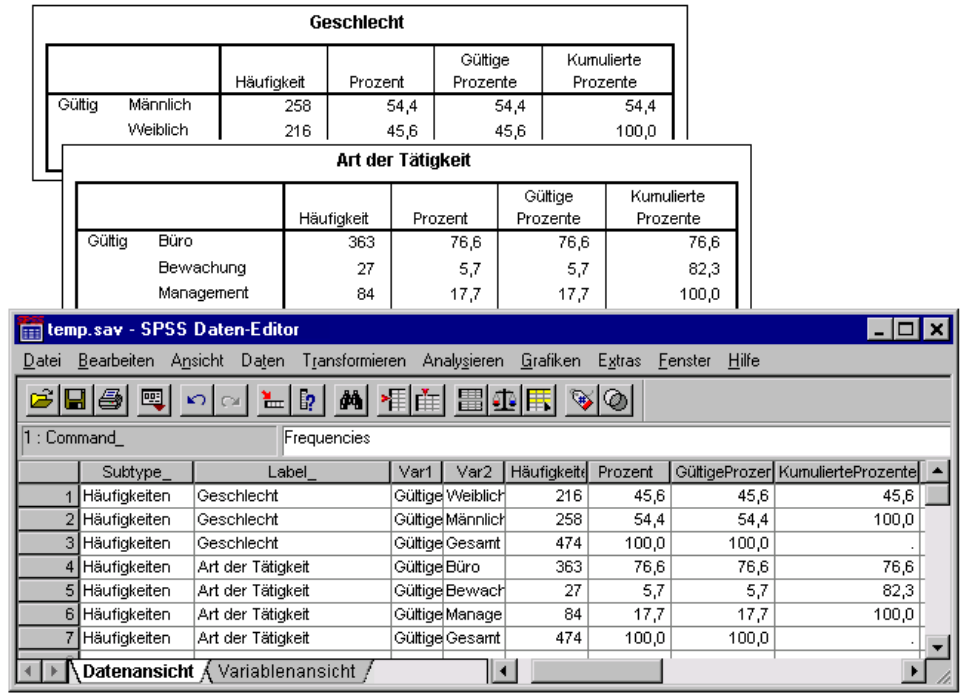
Werden mehrere Tabellen an dieselbe Datendatei weitergeleitet, wird jede Tabelle jeweils zur Datendatei hinzugefügt. Ähnlich wie beim Zusammenfügen von Datendateien werden hierbei die Fälle aus einer Datendatei in eine andere Datendatei aufgenommen (Menü “Daten”, “Dateien zusammenfügen”, “Fälle hinzufügen”).

- Bei jeder nachfolgenden Tabelle werden weitere Fälle zur Datendatei hinzugefügt.
- Weichen die Spaltenbeschriftungen in den Tabellen voneinander ab, können ggf. auch Variablen in die Datendatei aufgenommen werden. Bei Fällen aus anderen Tabellen, die keine Spalte mit der entsprechenden Beschriftung aufweisen, entstehen dabei fehlende Werte.

Beispiel: Mehrere Tabellen mit identischen Spaltenbeschriftungen

Mehrere Tabellen, die dieselben Spaltenbeschriftungen enthalten, resultieren in der Regel in sofort nutzbaren Datendateien; diese Dateien müssen nicht mehr nachbearbeitet werden. Die mit der Prozedur “Häufigkeit” erzeugten Häufigkeitstabellen besitzen beispielsweise immer dieselben Spaltenbeschriftungen.

Abbildung 48-9
Zwei Tabellen mit identischen Spaltenbeschriftungen



Geschlecht

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Männlich	258	54,4	54,4	54,4
	Weiblich	216	45,6	45,6	100,0

Art der Tätigkeit

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Büro	363	76,6	76,6	76,6
	Bewachung	27	5,7	5,7	82,3
	Management	84	17,7	17,7	100,0

	Subtype_	Label_	Var1	Var2	Häufigkeit	Prozent	GültigeProzer	KumulierteProzente
1	Häufigkeiten	Geschlecht	Gültige	Weiblich	216	45,6	45,6	45,6
2	Häufigkeiten	Geschlecht	Gültige	Männlich	258	54,4	54,4	100,0
3	Häufigkeiten	Geschlecht	Gültige	Gesamt	474	100,0	100,0	.
4	Häufigkeiten	Art der Tätigkeit	Gültige	Büro	363	76,6	76,6	76,6
5	Häufigkeiten	Art der Tätigkeit	Gültige	Bewach	27	5,7	5,7	82,3
6	Häufigkeiten	Art der Tätigkeit	Gültige	Manage	84	17,7	17,7	100,0
7	Häufigkeiten	Art der Tätigkeit	Gültige	Gesamt	474	100,0	100,0	.

- Die zweite Tabelle trägt weitere Fälle (Zeilen) zur Datendatei bei, führt jedoch keine neuen Variablen ein, weil die Spaltenbeschriftungen exakt übereinstimmen. So entstehen keine großen Bereiche mit fehlenden Daten.
- Die Werte für *Command_* und *Subtype_*+ sind identisch; der Wert für *Label_* zeigt die Quelltabelle für die einzelnen Fallgruppen, weil die beiden Häufigkeitstabellen einen anderen Titel besitzen.

Beispiel: Mehrere Tabellen mit verschiedenen Spaltenbeschriftungen

Für jede eindeutige Spaltenbeschriftung in den Tabellen, die an die Datendatei weitergeleitet werden, wird eine neue Variable in der Datendatei erstellt. Wenn die Tabellen unterschiedliche Spaltenbeschriftungen enthalten, führt dies zu Bereichen mit fehlenden Daten.

Abbildung 48-10
Zwei Tabellen mit unterschiedlichen Spaltenbeschriftungen

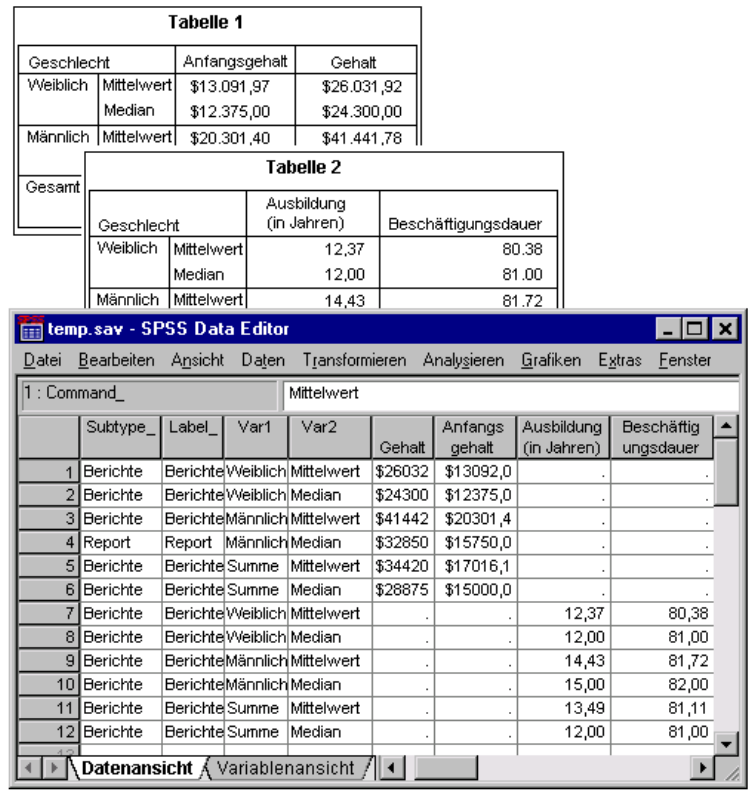


Tabelle 1

Geschlecht		Anfangsgehalt	Gehalt
Weiblich	Mittelwert	$13.091,97	$26.031,92
	Median	$12.375,00	$24.300,00
Männlich	Mittelwert	$20.301,40	$41.441,78
Gesamt			

Tabelle 2

Geschlecht		Ausbildung (in Jahren)	Beschäftigungsdauer
Weiblich	Mittelwert	12,37	80,38
	Median	12,00	81,00
Männlich	Mittelwert	14,43	81,72

temp.sav - SPSS Data Editor

Datei Bearbeiten Ansicht Daten Transformieren Analysieren Grafiken Extras Fenster

1 : Command_ | Mittelwert

	Subtype_	Label_	Var1	Var2	Gehalt	Anfangsgehalt	Ausbildung (in Jahren)	Beschäftigungsdauer
1	Berichte	Berichte	Weiblich	Mittelwert	$26032	$13092,0	.	.
2	Berichte	Berichte	Weiblich	Median	$24300	$12375,0	.	.
3	Berichte	Berichte	Männlich	Mittelwert	$41442	$20301,4	.	.
4	Report	Report	Männlich	Median	$32850	$15750,0	.	.
5	Berichte	Berichte	Summe	Mittelwert	$34420	$17016,1	.	.
6	Berichte	Berichte	Summe	Median	$28875	$15000,0	.	.
7	Berichte	Berichte	Weiblich	Mittelwert	.	.	12,37	80,38
8	Berichte	Berichte	Weiblich	Median	.	.	12,00	81,00
9	Berichte	Berichte	Männlich	Mittelwert	.	.	14,43	81,72
10	Berichte	Berichte	Männlich	Median	.	.	15,00	82,00
11	Berichte	Berichte	Summe	Mittelwert	.	.	13,49	81,11
12	Berichte	Berichte	Summe	Median	.	.	12,00	81,00

Datenansicht | Variablenansicht

- Die erste Tabelle enthält die Spalten *Anfangsgehalt* und *Aktuelles Gehalt*, die in der zweiten Tabelle nicht vorhanden sind. Bei Fällen aus der zweiten Tabelle entstehen so fehlende Werte für diese Variablen.
- Umgekehrt enthält die zweite Tabelle die Spalten *Schulabschluß* und *Beschäftigungsdauer*, die in der ersten Tabelle nicht vorhanden sind. Bei Fällen aus der ersten Tabelle entstehen entsprechend fehlende Werte für diese Variablen.
- Nicht übereinstimmende Variablen wie in diesem Beispiel können sogar bei Tabellen auftreten, die denselben Untertyp aufweisen. In diesem Beispiel besitzen beide Tabellen denselben Untertyp.

Beispiel: Datendateien, die nicht aus mehreren Tabellen erzeugt wurden

Falls eine Tabelle nicht dieselbe Anzahl an Zeilenelementen enthält wie die anderen Tabellen, wird keine Datendatei erstellt. Hierbei muß nicht die Anzahl der Zeilen identisch sein, sondern die Anzahl der Zeilen*elemente*, die zu Variablen in der Datendatei werden. Beispiel: Eine Kreuztabelle mit zwei Variablen und eine Kreuztabelle mit drei Variablen enthalten verschieden viele Zeilenelemente, weil die Variable "layer" in der Zeilenvariable in der Standard-Kreuztabellenanzeige mit drei Variablen verschachtelt ist.

Abbildung 48-11
Tabellen mit unterschiedlich vielen Zeilenelementen

Employment Category * Gender Crosstabulation

		Gender		
		Female	Male	Total
Employment Category	Clerical	206	157	363
	Custodial	0	27	27
	Manager	10	74	84
Total		216	258	474

Employment Category * Gender * Minority Classification Crosstabulation

Minority Classification			Gender		
			Female	Male	Total
No	Employment Category	Clerical	166	110	276
		Custodial	0	14	14
		Manager	10	70	80
	Total		176	194	370
Yes	Employment Category	Clerical	40	47	87
		Custodial	0	13	13
		Manager	0	4	4
	Total		40	64	104

Steuern von Spaltenelementen zum Steuern von Variablen in der Datendatei

Im Dialogfeld "Optionen" des Ausgabeverwaltungssystem-Bedienfelds können Sie angeben, welche Dimensionselemente sich in den Spalten befinden und daher zum Erstellen der Variablen in der erzeugten Datendatei verwendet werden sollen. Dies entspricht dem Pivotieren der Tabelle im Viewer.

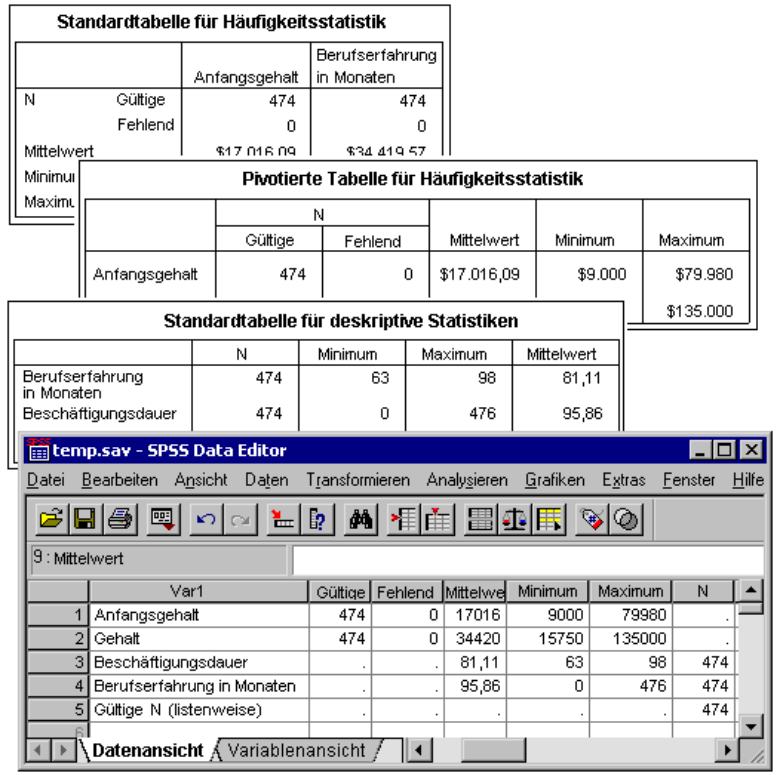
Beispiel: Bei der Prozedur "Häufigkeiten" wird eine deskriptive Statistiktabelle mit Statistiken in den Zeilen gebildet, bei der Prozedur "Deskriptive Statistiken" dagegen eine deskriptive Statistiktabelle mit Statistiken in den Zeilen. Wenn beide

Tabellentypen sinnvoll in dieselbe Datendatei aufgenommen werden sollen, muss die Spaltendimension für eine der Tabellentypen geändert werden.

Bei beiden Tabellentypen trägt die Statistikdimension den Elementnamen “Statistik”. Um die statistischen Werte aus der Häufigkeitstabelle in die Spalten zu übernehmen, reicht es daher, den Eintrag “Statistik” (in Anführungszeichen) zur Liste der Dimensionsnamen im Dialogfeld “Optionen” hinzuzufügen.

Abbildung 48-12
Dialogfeld “OMS: Optionen”

Abbildung 48-13
Kombinieren verschiedener Tabellentypen in einer Datendatei durch Pivotieren von Dimensionselementen



Standardtabelle für Häufigkeitsstatistik

		Anfangsgehalt	Berufserfahrung in Monaten
N	Gültige	474	474
	Fehlend	0	0
Mittelwert		$17.016,09	$34.419,57
Minimu			
Maximu			

Pivotierte Tabelle für Häufigkeitsstatistik

	N				
	Gültige	Fehlend	Mittelwert	Minimum	Maximum
Anfangsgehalt	474	0	$17.016,09	$9.000	$79.980
					$135.000

Standardtabelle für deskriptive Statistiken

	N	Minimum	Maximum	Mittelwert
Berufserfahrung in Monaten	474	63	98	81,11
Beschäftigungsdauer	474	0	476	95,86

temp.sav - SPSS Data Editor

Datei Bearbeiten Ansicht Daten Transformieren Analysieren Grafiken Extras Fenster Hilfe

9 : Mittelwert

	Var1	Gültige	Fehlend	Mittelwe	Minimum	Maximum	N
1	Anfangsgehalt	474	0	17016	9000	79980	.
2	Gehalt	474	0	34420	15750	135000	.
3	Beschäftigungsdauer	.	.	81,11	63	98	474
4	Berufserfahrung in Monaten	.	.	95,86	0	476	474
5	Gültige N (listenweise)	.	.	.	.	.	474

Datenansicht Variablenansicht

Bei einigen Variablen sind fehlende Werte vorhanden, weil die Tabellenstrukturen immer noch nicht völlig identisch mit den Statistiken in den Spalten sind.

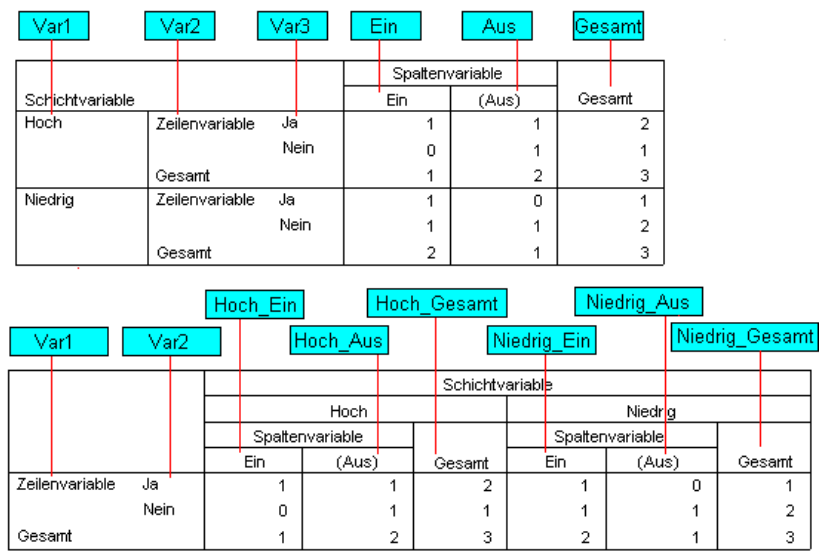
Variablennamen in Datendateien aus dem OMS

Das OMS erzeugt gültige, eindeutige Variablennamen aus den Spaltenbeschriftungen:

- Den Zeilen- und Schichtenelemeten werden generische Variablennamen zugewiesen. Diese bestehen aus dem Präfix *Var* und einer laufenden Nummer.

- Zeichen, die in Variablennamen nicht zulässig sind (z. B. Leerzeichen, Klammern usw.) werden entfernt. “Diese (Spalten-) Beschriftung” würde beispielsweise zu einer Variable mit der Bezeichnung *DieseSpalten-Beschriftung*.
- Ist das erste Zeichen der Beschriftung zwar grundsätzlich in Variablennamen zulässig, nicht jedoch als erstes Zeichen (z. B. Ziffern), wird das Symbol “@” als Präfix vorangestellt. “2tes” würde beispielsweise zu einer Variablen mit der Bezeichnung *@2tes*.
- Unterstriche und Punkte am Ende von Beschriftungen werden aus den resultierenden Variablennamen entfernt. (Die Unterstriche am Ende der automatisch erzeugten Variablen *Command_*, *Subtype_* und *Label_* bleiben erhalten.)
- Enthält die Spaltendimension mehrere Elemente, werden die Variablennamen aus einer Kombination der Kategorienbeschriftungen gebildet; die Kategorienbeschriftungen werden dabei durch einen Unterstrich getrennt. Gruppenbeschriftungen werden nicht berücksichtigt. Wenn beispielsweise *VarB* in *VarA* in den Spalten verschachtelt ist, erhalten Sie Variablen wie *CatA1_CatB1*, nicht jedoch *VarA_CatA1_VarB_CatB1*.

Abbildung 48-14
Variablennamen, die aus Tabellenelementen gebildet wurden



OXML-Tabellenstruktur

Output XML (OXML) ist XML, das dem Schema *spss-output* entspricht. Eine ausführliche Beschreibung des Schemas finden Sie in der Datei *SPSSOutputXML_schema.htm* im Ordner *help\main* im SPSS-Installationsordner.

- Die Befehls- und Untertyp-IDs im OMS dienen als Werte für die Attribute command und subType in OXML. Ein Beispiel lautet folgendermaßen:

```
<command text="Häufigkeiten" command="Häufigkeiten"...>
  <pivotTable text="Geschlecht" label="Geschlecht" subType="Häufigkeiten"...>
```

- Die Ausgabesprache sowie die Einstellungen für die Anzeige von Variablennamen/Beschriftungen und Werte/Wertelabels wirken sich nicht auf die Werte der OMS-Attribute command und subType aus.
- Bei XML wird zwischen Groß- und Kleinschreibung unterschieden. Der subType-Attributwert "häufigkeiten" ist *nicht* identisch mit dem subType-Attributwert "Häufigkeiten".
- Alle in einer Tabelle angezeigten Informationen befinden sich in Attributwerten in OXML. Auf der Ebene einzelner Zellen besteht OXML aus "leeren" Elementen, die zwar Attribute enthalten, jedoch keine "Inhalte" (außer den Inhalten in den Attributwerten).
- Die Tabellenstruktur in OMXL wird zeilenweise dargestellt. Die Spaltenelemente sind in den Zeilen verschachtelt, und einzelne Zellen sind wiederum in den Spaltenelementen verschachtelt.

```
<pivotTable...>
 <dimension axis='row'...>
  <dimension axis='column'...>
   <category...>
    <cell text='...' number='...' decimals='...'/>
   </category>
   <category...>
    <cell text='...' number='...' decimals='...'/>
   </category>
  </dimension>
</dimension>
   ...
</pivotTable>
```

Das vorausgehende Beispiel ist eine vereinfachte Darstellung der Struktur, die die Beziehungen zwischen Nachfolgern und Vorgängern veranschaulicht. Das Beispiel zeigt jedoch nicht notwendigerweise die direkt über- oder untergeordneten Elemente, weil in der Regel verschachtelte Ebenen von Elementen vorliegen.

Die nachstehende Abbildung zeigt eine einfache Häufigkeitstabelle und die vollständige XML-Ausgabedarstellung dieser Tabelle.

Abbildung 48-15
Einfache Häufigkeitstabelle

Geschlecht

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Männlich	258	54,4	54,4	54,4
	Weiblich	216	45,6	45,6	100,0
	Gesamt	474	100,0	100,0	

Abbildung 48-16
XML-Ausgabe für eine einfache Häufigkeitstabelle

```
<?xml version="1.0" encoding="UTF-8" ?>
<outputTreeoutputTree xmlns="http://xml.spss.com/spss/oms"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:schemaLocation="http://xml.spss.com/spss/oms
 http://xml.spss.com/spss/oms/spss-output-1.0.xsd">
<command text="Häufigkeiten" command="Häufigkeiten"
 displayTableValues="label" displayOutlineValues="label"
 displayTableVariables="label" displayOutlineVariables="label">
 <pivotTable text="Geschlecht" label="Geschlecht" subType="Häufigkeiten"
  varName="geschl" variable="true">
  <dimension axis="row" text="Geschlecht" label="Geschlecht"
   varName="geschl" variable="true">
   <group text="Gültig">
    <group hide="true" text="Dummy">
     <category text="Weiblich" label="Weiblich" string="w"
      varName="geschl">
      <dimension axis="column" text="Statistik">
       <category text="Häufigkeit">
        <cell text="216" number="216"/>
       </category>
       <category text="Prozent">
        <cell text="45,6" number="45,569620253165" decimals="1"/>
       </category>
       <category text="Gültige Prozente">
```

```
          <cell text="45,6" number="45,569620253165" decimals="1"/>
         </category>
         <category text="Kumulierte Prozente">
          <cell text="45,6" number="45,569620253165" decimals="1"/>
         </category>
        </dimension>
       </category>
       <category text="Männlich" label="Männlich" string="m" varName="geschl">
        <dimension axis="column" text="Statistik">
         <category text="Häufigkeit">
          <cell text="258" number="258"/>
         </category>
         <category text="Prozent">
          <cell text="54,4" number="54,430379746835" decimals="1"/>
         </category>
         <category text="Gültige Prozente">
          <cell text="54,4" number="54,430379746835" decimals="1"/>
         </category>
         <category text="Kumulierte Prozente">
          <cell text="100,0" number="100" decimals="1"/>
         </category>
        </dimension>
       </category>
      </group>
      <category text="Gesamt">
       <dimension axis="column" text="Statistik">
        <category text="Häufigkeit">
         <cell text="474" number="474"/>
        </category>
        <category text="Prozent">
         <cell text="100,0" number="100" decimals="1"/>
        </category>
        <category text="Gültige Prozente">
         <cell text="100,0" number="100" decimals="1"/>
        </category>
       </dimension>
      </category>
     </group>
    </dimension>
   </pivotTable>
  </command>
</outputTree>
```

Eine einfache, kleine Tabelle kann zu beträchtlichen Mengen an XML führen. Der Grund hierfür liegt teilweise darin, dass die XML einige Informationen enthält, die aus der ursprünglichen Tabelle nicht ohne weiteres ersichtlich sind, sowie einige Informationen, die in der ursprünglichen Tabelle nicht einmal vorlagen. Auch eine gewisse Redundanz ist vorhanden.

- Der Tabelleninhalt, wie er im Viewer in einer Pivot-Tabelle dargestellt wird (oder würde), ist in Textattributen enthalten. Ein Beispiel lautet folgendermaßen:

```
<command text="Häufigkeiten" command="Häufigkeiten"...>
```

- Die Ausgabesprache sowie die Einstellungen für die Anzeige von Variablennamen/Beschriftungen und Werte/Wertelabels wirken sich ggf. auf die Textattribute aus. In diesem Beispiel ist der Wert des Attributs text abhängig von der Ausgabesprache, der Wert des Attributs "command" bleibt dagegen unabhängig von der Ausgabesprache immer gleich.
- An allen Stellen, an denen Variablen oder Werte für Variablen in Zeilen- oder Spaltenbeschriftungen auftreten, enthält die XML ein Attribut text sowie mindestens einen weiteren Attributwert. Ein Beispiel lautet folgendermaßen:

```
<dimension axis="row" text="Geschlecht" label="Geschlecht" varName="geschl">
  ...<category text="Weiblich" label="Weiblich" string="w" varName="geschl">
```

- Bei einer numerischen Variable würde entsprechend ein Attribut number anstelle eines Attributs string verwendet. Das Attribut label ist nur dann vorhanden, wenn eine Beschriftung für die Variable oder die Werte definiert wurde.
- Die Elemente <cell> mit den Zellwerten für Zahlen enthalten das Attribut text sowie mindestens einen weiteren Attributwert. Ein Beispiel lautet folgendermaßen:

```
<cell text="45,6" number="45,569620253165" decimals="1"/>
```

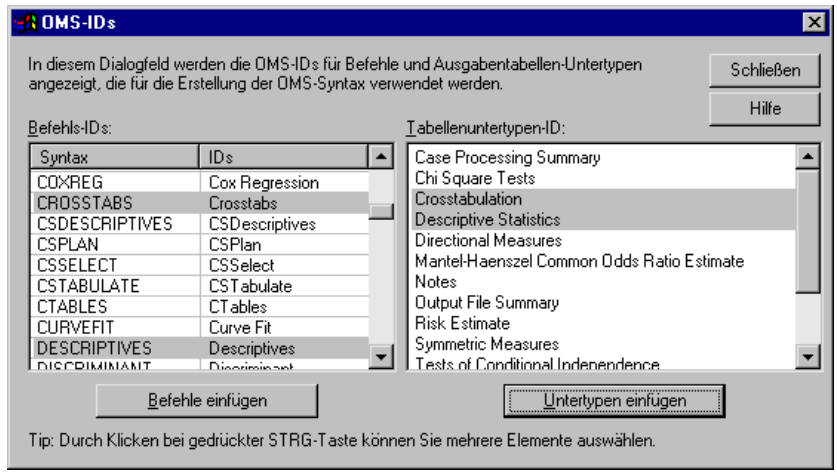
Das Attribut number ist der eigentliche, nicht gerundete numerische Wert, und das Attribut decimals bezeichnet die Anzahl der Dezimalstellen, die in der Tabelle angezeigt werden.

- Da die Spalten in den Zeilen verschachtelt sind, wird das Kategorieelement für die einzelnen Spalten für jede Zeile wiederholt. Beispiel: Die Statistik wird in den Spalten angezeigt, und daher wird das Element <category text="Häufigkeit"> dreimal in der XML aufgeführt: einmal für die Zeile "Männlich", einmal für die Zeile "Weiblich" und einmal für die Zeile "Gesamt".

OMS-IDs

Das Dialogfeld “OMS-IDs” soll Sie beim Schreiben von OMS-Befehlssyntax unterstützen. Mit diesem Dialogfeld können Sie ausgewählte Befehls- und Untertypen-IDs in ein Befehlssyntax-Fenster einfügen.

Abbildung 48-17
Dialogfeld “OMS-IDs”



So verwenden Sie das Dialogfeld “OMS-IDs”

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Extras
  OMS-IDs...

- Wählen Sie einen oder mehrere Befehls- oder Untertypen-IDs aus. (Halten Sie beim Klicken mit der Maus gleichzeitig die Strg-Taste gedrückt, um mehrere IDs in den einzelnen Listen auszuwählen.)

- Klicken Sie auf Befehle einfügen und/oder Untertypen einfügen.
  - Die Liste der verfügbaren Untertypen hängt von den zum jeweiligen Zeitpunkt ausgewählten Befehlen ab. Wenn mehrere Befehle ausgewählt wurden, stellt die Liste der verfügbaren Untertypen die Gesamtheit aller Untertypen dar, die für die jeweils ausgewählten Befehle verfügbar sind. Wenn keine Befehle ausgewählt wurden, werden alle Untertypen aufgeführt.

- Die IDs werden in das Hauptfenster für die Befehlssyntax an der jeweiligen Cursorposition eingefügt. Wenn keine Befehlssyntax-Fenster geöffnet sind, wird automatisch ein neues Syntax-Fenster geöffnet.
- Jede Befehls- und/oder Untertypen-ID wird beim Einfügen in Anführungszeichen eingeschlossen, da diese Anführungszeichen für die Befehlssyntax von OMS erforderlich sind.
- ID-Listen für die Schlüsselwörter COMMANDS und SUBTYPES müssen in Klammern eingeschlossen sein, wie in folgendem Beispiel:

```
/IF COMMANDS=['Kreuztabellen' 'Deskriptive Statistiken']
    SUBTYPES=['Kreuztabelle' 'Deskriptive Statistiken']
```

Kopieren von OMS-IDs aus Viewer-Gliederung

Sie können OMS-Befehls-IDs und OMS-Untertypen-IDs aus dem Viewer-Gliederungsfenster kopieren und einfügen.

▶ Klicken Sie im Gliederungsfenster mit der rechten Maustaste auf den Gliederungseintrag für das Element.

▶ Wählen Sie die Option OMS-Befehls-ID kopieren oder die Option OMS-Tabellenuntertyp kopieren.

Diese Methode unterscheidet sich in einem Punkt von der Verwendung des Dialogfelds “OMS-IDs”: Die kopierte ID wird nicht automatisch in das Befehlssyntax-Fenster eingefügt. Die ID wird einfach in die Zwischenablage kopiert, und Sie können sie anschließend an jeder gewünschten Stelle einfügen. Da die Werte für die Befehls- und die Untertypen-IDs exakt mit den zugehörigen Befehls- und Untertypen-Attributwerten in der XML-Ausgabe (OXML) übereinstimmen, ist diese Methode des Kopierens und Einfügens besonders hilfreich beim Schreiben von XSLT-Transformationen.

Kopieren von OMS-Labels

Statt der IDs können Sie Labels für die Verwendung mit dem Schlüsselwort LABELS kopieren. Labels können verwendet werden, um zwischen mehreren Diagrammen oder mehreren Tabellen desselben Typs zu unterscheiden, bei denen der Gliederungstext ein Attribut des jeweiligen Ausgabeobjekts angibt, beispielsweise

die Variablennamen oder -labels. Es gibt jedoch eine Reihe von Faktoren, die den Labeltext beeinflussen können:

- Wenn die Verarbeitung aufgeteilter Dateien aktiviert ist, kann die Gruppen-ID für die aufgeteilte Datei an die Beschriftung angehängt werden.
- Labels, die Informationen über Variablen oder Werte enthalten, hängen unter anderem von den Einstellungen für die Anzeige von Variablennamen/-werten und den Werten/Wertelabels im Gliederungsfenster ab (Menü “Bearbeiten”, “Optionen”, Registerkarte “Beschriftung der Ausgabe”).
- Labels richten sich außerdem nach der aktuellen Einstellung für die Ausgabesprache (Menü “Bearbeiten”, “Optionen”, Registerkarte “Allgemein”).

So kopieren Sie OMS-Beschriftungen

▶ Klicken Sie im Gliederungsfenster mit der rechten Maustaste auf den Gliederungseintrag für das Element.

▶ Wählen Sie OMS-Label kopieren aus.

Wie bei den Befehls- und Untertypen-IDs müssen die Beschriftungen in Anführungszeichen und die gesamte Liste in eckige Klammern eingeschlossen sein, wie in folgendem Beispiel:

```
/IF LABELS=['Art der Tätigkeit' 'Schulabschluß']
```

Anhang

A

Administrator für Datenbankzugriff

Der Administrator für Datenbankzugriff ist ein Dienstprogramm, das zum Vereinfachen des Zugriffs auf große oder komplexe Datenquellen mit dem Datenbank-Assistenten dient. Mit dem Administrator für Datenbankzugriff können Benutzer und Administratoren folgende Anpassungen der Datenquellen vornehmen:

- Erstellen von Aliasen für Datenbanktabellen und Datenbankfelder.
- Erstellen von Variablennamen für Felder.
- Ausblenden unwesentlicher Tabellen und Felder.

Mit dem Administrator für Datenbankzugriff wird Ihre Datenbank nicht wirklich geändert. Stattdessen werden Dateien erzeugt, die Ihre Informationen enthalten. Diese Dateien fungieren als *Ansichten* für Datenbanken.

Mit dem Administrator für Datenbankzugriff können Sie bis zu drei verschiedene Ansichten pro Datenbank festlegen: Unternehmensebene, Abteilungsebene und persönliche Ebene. Sowohl mit dem Administrator für Datenbankzugriff als auch mit dem Datenbank-Assistenten werden solche Dateien an folgenden Namen erkannt:

- Unternehmensebene: *dba01.inf*
- Abteilungsebene: *dba02.inf*
- Persönliche Ebene: *dba03.inf*

Jede Datei enthält für die jeweilige Ebene spezifische Informationen über die Anzahl von Datenquellen. Die Datei *dba03.inf* könnte beispielsweise Informationen über die persönliche Ebene für eine Kontendatenbank eines Unternehmens sowie für die Arbeitszeitendatenbank ihrer Firma und für die Datenbank enthalten, die Ihre CD-Sammlung dokumentiert.

Wenn Sie den Administrator für Datenbankzugriff aufrufen, wird Ihr Systempfad nach solchen Dateien durchsucht. Es werden Ihnen automatisch Informationen zu allen drei Ansichten einer beliebigen Datenquelle angezeigt, die Sie konfiguriert haben.

Vererbung und Prioritäten. Der Datenbank-Assistent zeigt für die Datenquelle die Ansicht der niedrigsten Ebene, die im Systempfad gefunden wird. Hierbei stellen Unternehmen, Abteilung und persönlich die Ebenen in absteigender Reihenfolge dar. Jede Datei für eine Ebene enthält Informationen, die sämtliche Datenquellen dieser Ebene betreffen. So verfügt beispielsweise die Marketing-Abteilung über eine Datei namens *dba02.inf* mit den Alias-Informationen für alle Datenbankansichten, die für die Marketing-Abteilung angelegt wurden. Jeder Mitarbeiter in der Marketing-Abteilung verfügt über eine Datei mit der Bezeichnung *dba03.inf*, die die benutzerdefinierten Ansichten aller von diesem Mitarbeiter verwendeten Datenbanken enthält.

Im Administrator für Datenbankzugriff werden Aliase, Variablennamen und Ausblendereihenfolgen von oben nach unten vererbt.

Beispiele:

- Wenn Sie eine Regionstabelle auf der Unternehmensebene ausblenden, wird sie gleichermaßen auf der Abteilungs- und persönlichen Ebene ausgeblendet. Die Tabelle wird dann im Datenbank-Assistenten nicht angezeigt.
- Vom Feld *JOBKAT* in der Tabelle "MitarbeiterUmsatz" wird auf der Unternehmensebene kein Alias erstellt. Es wird hingegen ein Alias als "Jobkategorien" auf der Abteilungsebene erstellt. Es erscheint als "Jobkategorien" auf der persönlichen Ebene. Wenn Sie außerdem auf der persönlichen Ebene ein Alias "Mitarbeiterinformation" von der Tabelle "MitarbeiterUmsatz" erstellen, wird das Originalfeld (*MitarbeiterUmsatz.JOBKAT*) im Datenbank-Assistenten als *'Mitarbeiterinformation'.'Jobkategorien'* angezeigt.

Um den Administrator für Datenbankzugriff zu starten, führen Sie die Datei *spssdbca.exe* aus. Diese Datei befindet sich im Installationsverzeichnis von SPSS. Weitere Informationen erhalten Sie in der Online-Hilfe zum Administrator für Datenbankzugriff.

Anhang

B

Anpassen von HTML-Dokumenten

Sie können benutzerdefinierten HTML-Code automatisch in die Dokumente einfügen lassen, die in das HTML-Format exportiert werden sollen. Zu diesen Code-Einträgen gehören unter anderem:

- Titel von HTML-Dokumenten,
- Angaben zum Dokumenttyp,
- Meta-Tags und Skript-Code, wie zum Beispiel JavaScript,
- Text, der vor und nach der Ausgabe angezeigt wird.

So fügen Sie benutzerdefinierten HTML-Code in exportierte Viewer-Dokumente ein:

▶ Öffnen Sie die Datei *htmlfram.txt*, die sich in Ihrem SPSS-Verzeichnis befindet, in einem Text-Editor.

▶ Ersetzen Sie nun die in den "Feldern" stehenden Kommentare durch den Text oder HTML-Code, den Sie in Ihre exportierten HTML-Dokumente einfügen möchten. Die "Felder" sind mit doppelten geöffneten Klammern (<<) gekennzeichnet.

▶ Speichern Sie die Datei im Format "Textdatei".

Anmerkung: Wenn Sie den Namen oder den Speicherort der Textdatei ändern, diese aber auch weiterhin zum Anpassen Ihrer exportierten HTML-Ausgabe verwenden möchten, müssen Sie in der Systemregistrierung die entsprechenden Änderungen vornehmen.

Inhalt und Format einer Textdatei für benutzerdefinierten HTML-Code

Wenn Sie HTML-Code automatisch in Ihre HTML-Dokumente einfügen lassen möchten, müssen Sie diesen zunächst in eine einfache Textdatei eingeben. Dazu stehen Ihnen in der Textdatei sechs Felder zur Verfügung, die jeweils durch zwei geöffnete spitze Klammern (<<) in der vorangehenden Zeile gekennzeichnet sind:

<<
Text oder Code, der im Dokumentkopf noch vor dem eigentlichen `<HTML>`-Dokument eingefügt werden soll (beispielsweise Kommentare mit Angaben zum Dokumenttyp),

<<
Text für den Titel des Dokuments (dieser wird in der Titelleiste angezeigt),

<<
Meta-Tags oder Skript-Code (beispielsweise JavaScript),

<<
HTML-Code für Optionen im `<BODY>`-Tag (beispielsweise Code für die Hintergrundfarbe),

<<
Text und/oder HTML-Code, der nach der exportierten Ausgabe eingefügt wird (beispielsweise Informationen zum Urheberrecht),

<<
Text und/oder HTML-Code, der vor der exportierten Ausgabe eingefügt wird (beispielsweise Firmenname, Logo usw.).

So verwenden Sie eine andere Datei oder einen anderen Speicherort für den benutzerdefinierten HTML-Code:

Wenn Sie den Namen oder Speicherort von *htmlfram.txt* ändern, die Datei aber auch weiterhin für die benutzerdefinierte HTML-Ausgabe verwenden möchten, müssen Sie in der Systemregistrierung die nötigen Änderungen vornehmen.

- Wählen Sie im Menü "Start" von Windows die Option Ausführen. Geben Sie anschließend den Befehl regedit ein, und klicken Sie auf OK.

- Wählen Sie im linken Fenster des Registrierungseditors die folgenden Optionen:

```
HHKEY_CURRENT_USER
  Software
    SPSS
      SPSS für Windows
        14.0
          Spsswin
```

- Doppelklicken Sie im rechten Fensterbereich auf die Zeichenfolge HTMLFormatFile.
- Geben Sie unter Wert den vollständigen Pfad und Namen der Textdatei an, in der die benutzerdefinierten HTML-Spezifikationen gespeichert sind (beispielsweise *c:\myfiles\htmlstuf.txt*).

Beispieltextdatei für benutzerdefinierten HTML-Code

```
<<
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 3.2//EN">
<<
NVI, Inc.
<<
<META NAME="keywords" CONTENT="Geräte, Apparate, Krimskrams">
<<
bgcolor="#FFFFFF"
<<
<H4 align=center>Diese Seite wird präsentiert von ...
<br><br>
<IMG SRC="spss2.gif" align=center></H4>
<<
<h2 align=center>NVI Sales</h2>
<h3 align=center>Regionale Umsatzdaten</h3>
```

Beispiel-HTML-Quelltext für benutzerdefinierten HTML-Code

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 3.2//EN">
<HTML>
<HEAD>
<TITLE>
NVI Sales, Inc.
```

```
</TITLE>
<META NAME="keywords" CONTENT="Geräte, Apparate, Krimskrams">
</HEAD>
<BODY bgcolor="#FFFFFF">
<h2 align=center>NVI Sales</h2>
<h3 align=center>Regionale Umsatzdaten</h3>

[Exportierte Ausgabe]

<H4 align=center>Diese Seite wird präsentiert von ...
<br><br>
<IMG SRC="spss2.gif" align=center></H4>
</BODY>
</HTML>
```

Index

Abbrechen (Schaltfläche), 11
Abhängiger *T*-Test
  in T-Test bei gepaarten Stichproben, 425
Abweichungskontraste
  in GLM, 450
Access (Microsoft), 31
ActiveX-Objekte, 282
Aggregieren von Daten, 232
  Aggregierungsfunktionen, 235
  Variablennamen und -labels, 236
Ähnlichkeiten
  in der hierarchischen Clusteranalyse, 543
Ähnlichkeitsmaße
  in der hierarchischen Clusteranalyse, 547
  in Distanzen, 478
Aktive Datei, 78, 81
  Erstellen einer temporären aktiven Datei, 81
  in Zwischenspeicher ablegen, 81
  virtuelle aktive Datei, 78
Aktives Fenster, 8
Algorithmen, 17
Alpha-Faktorisierung, 524
Alpha-Koeffizient
  in der Reliabilitätsanalyse, 621, 623
Analyse von Mehrfachantworten
  Häufigkeitstabellen, 596
  Kreuztabelle, 598
  Mehrfachantworten: Häufigkeiten, 596
  Mehrfachantworten: Kreuztabellen, 598
Anderson-Rubin-Faktorwerte, 528
Andrew-Wellen-Schätzer
  in der Explorativen Datenanalyse, 381
Anfänglicher Schwellenwert
  in der Two-Step-Clusteranalyse, 537
Anlegen von Lesezeichen für Ansichten von Pivot-Tabellen, 326
Anmelden bei einem Server, 85
ANOVA
  in einfaktorieller ANOVA, 433
  in GLM - Univariat, 443
  in Mittelwerte, 410
  Modell, 447
Anteilsschätzungen
  beim Bilden der Rangfolge, 182
Anzahl der Fälle
  in Mittelwerte, 410
  in OLAP-Würfel, 416
  in Zusammenfassen, 403
Anzeigeformate, 103
Anzeigereihenfolge, 319–320
Arbeitsspeicher, 666
Auflisten von Fällen, 399
Aufteilen von Tabellen, 350
  Festlegen von Tabellenumbrüchen, 350
Ausblenden, 273–274, 328–329, 689
  Dimensionsbeschriftungen, 329
  Erklärungen, 329
  Fußnoten, 329
  Prozedurergebnisse, 274
  Symbolleisten, 689
  Titel, 329
  Zeilen und Spalten, 328

Ausgabe, 271, 273–275, 280–281, 284–285, 305, 307, 348, 668, 698
ändern, 348
Ändern der Ausgabesprache, 666
Ausblenden, 273
Ausrichtung, 275, 668
Einblenden, 273
Einfügen in andere Anwendungen, 280
exportieren, 285, 698
Kopieren, 274–275
kopieren in andere Anwendungen, 281
Kopieren und Einfügen mehrerer Objekte, 284
löschen, 274
speichern, 305
Text, 307
Verschieben, 274
Viewer, 271
zentrieren, 275, 668
Ausgabeobjekttypen
bei OMS, 751
Ausgabeverwaltungssystem (OMS), 747, 778
Ausgeschlossene Fälle, 712
auswerten, 712
Ausgeschlossene Residuen
in GLM, 456
in Lineare Regression, 489
Auslöseereignisse, 723
Autoskripts, 723
Ausreißer
in der Explorativen Datenanalyse, 381
in der Two-Step-Clusteranalyse, 537
in Lineare Regression, 487
Ausrichtung, 108, 275, 344, 668
Ausgabe, 275, 668
im Daten-Editor, 108
in Zellen, 344
Ausschließen der Ausgabe aus dem Viewer bei OMS, 763
Auswählen von Fällen, 238
auf der Grundlage von Auswahlkriterien, 240
Bereich von Fällen, 242
Datumsbereich, 242
Zeitbereich, 242
Zufallsstichprobe, 241
Auswahlmethoden, 348
Auswählen von Zeilen und Spalten in Pivot-Tabellen, 348
Auswahlvariable
in Lineare Regression, 487
automatisierte Produktion, 695
Automatisierungsobjekte, 725, 727–728, 730–731
Eigenschaften, 730
Konventionen für die Benennung von Variablen, 727
Methoden, 730
Objektkatalog, 731
Typen, 727
verwenden in Skripts, 725, 728, 731
Autoskripts, 684, 720, 723
Auslöseereignisse, 723
Autoskript-Datei, 724
erstellen, 720

Balkendiagramme
in Häufigkeiten, 368
Bartlett-Faktorwerte, 528
Bartlett-Test auf Sphärizität
in der Faktorenanalyse, 523
Baumtiefe
in der Two-Step-Clusteranalyse, 537
Bearbeiten von Daten, 113–114
Bedingte Transformationen, 167

Befehls-IDs, 753
Befehlssprache, 353
Befehlssyntax, 353, 361, 666, 668–669, 687, 692, 695
  ausführen, 361
  Ausführen mit Symbolleisten-Schaltflächen, 692
  Ausgabe-Log, 357
  einfügen, 356
  Formatierungen im Produktionsmodus, 706
  Hinzufügen zu Menüs, 687
  Journaldatei, 359, 362, 666
  Log, 668–669
  Regeln für den SPSS-Produktionsmodus, 695
  Syntaxregeln, 353
  Zugreifen auf *SPSS Command Syntax Reference*, 17
Befehlszeilenschalter, 707
  SPSS-Produktionsmodus, 707
Benutzerdefinierte fehlende Werte, 106
Benutzerdefinierte Modelle
  in GLM, 447
Beobachtete Anzahl
  in Kreuztabellen, 395
Beobachtete Mittelwerte
  in GLM - Univariat, 458
Berechnen von Variablen, 165
  Berechnen von neuen String-Variablen, 168
Berechtigungen, 93
Bereich
  in Deskriptive Statistiken, 373
  in Häufigkeiten, 366
  in Mittelwerte, 410
  in OLAP-Würfel, 416
  in Verhältnisstatistiken, 637
  in Zusammenfassen, 403
Bereichseinteilung von Datenwerten, 152
Bericht in Spalten, 613
  fehlende Werte, 618
  Gesamtergebnis, 618
  Gesamtergebnisspalten, 616
  Seiteneinstellung, 617
  Seitenformat, 610
  Seitennumerierung, 618
  Spaltenformat, 607
  zusätzliche Funktionen beim Befehl, 619
  Zwischenergebnisse, 617
Bericht in Zeilen, 605
  Break-Abstand, 609
  Break-Spalten, 605
  Datenspalten, 605
  fehlende Werte, 610
  Fußzeilen, 612
  Seiteneinstellung, 609
  Seitenformat, 610
  Seitennumerierung, 610
  Sortierfolgen, 605
  Spaltenformat, 607
  Titel, 612
  Variablen in Titel, 612
  zusätzliche Funktionen beim Befehl, 619
Berichte
  Berichte in Spalten, 613
  Berichte in Zeilen, 605
  Dividieren von Spaltenwerten, 616
  Gesamtergebnisspalten, 616
  Multiplizieren von Spaltenwerten, 616
  Vergleichen von Spalten, 616
  zusammengesetzte Gesamtergebnisse, 616
Berichte in Spalten, 613
Beschriftungen, 320–321
  Einfügen von Gruppenbeschriftungen, 320
  im Vergleich mit Untertypennamen in OMS, 755

löschen, 321
Beta-Koeffizienten
in Lineare Regression, 493
Bewertung, 214
Anzeigen der geladenen Modelle, 218
Laden von gespeicherten Modellen, 215
Bilden der Rangfolge, 181
gebundene Werte, 183
Perzentile, 182
relative Ränge, 182
Savage-Werte, 182
Bivariate Korrelationen
Fehlende Werte, 466
Korrelationskoeffizienten, 463
Optionen, 466
Signifikanzniveau, 463
Statistiken, 466
zusätzliche Funktionen beim Befehl, 467
Block-Distanz
in Distanzen, 477
Blom-Schätzungen, 182
BMP-Dateien, 285, 292–293, 698
Exportieren von Diagrammen, 285, 292–293, 698
Bonferroni
in einfaktorieller ANOVA, 437
in GLM, 453
Box-M-Test
in der Diskriminanzanalyse, 509
Boxplots
in der Explorativen Datenanalyse, 382
Vergleichen von Faktorstufen, 382
Vergleichen von Variablen, 382
Break-Variablen
in Aggregieren von Daten, 232
Brown-Forsythe-Statistik
in einfaktorieller ANOVA, 440

C nach Dunnett
in einfaktorieller ANOVA, 437
in GLM, 453
Chi-Quadrat, 562
auf Unabhängigkeit, 391
erwartete Werte, 564
erwarteter Bereich, 564
Exakter Test nach Fisher, 391
fehlende Werte, 565
in Kreuztabellen, 391
Kontinuitätskorrektur nach Yates, 391
Likelihood-Quotient, 391
Optionen, 565
Pearson-Korrelationskoeffizient, 391
Statistiken, 565
Test bei einer Stichprobe, 562
Zusammenhang linear-mit-linear, 391
Chi-Quadrat-Distanz
in Distanzen, 477
Cluster-Häufigkeiten
in der Two-Step-Clusteranalyse, 541
Clusteranalyse
Auswählen einer Prozedur, 531
Clusterzentrenanalyse, 551
Effizienz, 556
Hierarchische Clusteranalyse, 543
Clusterzentrenanalyse, 551
Beispiele, 551
Cluster-Zugehörigkeit, 558
Distanzen der Cluster, 558
Effizienz, 556
fehlende Werte, 558
Iterationen, 557

Konvergenzkriterien, 557
Methoden, 551
Speichern von Cluster-Informationen, 558
Statistiken, 551, 558
Übersicht, 551
zusätzliche Funktionen beim Befehl, 559
Cochran-*Q*
in Tests bei mehreren verbundenen Stichproben, 590
Cochran-Statistik
in Kreuztabellen, 391
Cohen-Kappa
in Kreuztabellen, 391
Cook-Distanz
in GLM, 456
in Lineare Regression, 489
Cramér-*V*
in Kreuztabellen, 391
Cronbachs Alpha
in der Reliabilitätsanalyse, 621, 623

d
in Kreuztabellen, 391
Data List (Befehlssyntax), 78
Vergleich mit 'Get Data', 78
Dateien, 280
Hinzufügen von Textdateien im Viewer, 280
öffnen, 25
Dateitransformationen, 245–246
Aggregieren von Daten, 232
Gewichten von Fällen, 243
Sortieren von Fällen, 222
Transponieren von Variablen und Fällen, 223
Umstrukturieren von Daten, 245–246
Verarbeitung von aufgeteilten Dateien, 237
Zusammenfügen von Datendateien, 224, 229
Daten aus Dimensions, 62
Daten im CSV-Format, 50
Daten-Editor, 95, 98, 107–108, 110–118, 120, 687
Ändern des Datentyps, 116
Anzeigeoptionen, 118
Ausrichtung, 108
Bearbeiten von Daten, 113–114
Datenansicht, 96
Definieren von Variablen, 98
Drucken, 120
Einfügen von neuen Fällen, 115
Einfügen von neuen Variablen, 115
Eingeben von Daten, 110
Eingeben von nichtnumerischen Daten, 111
Eingeben von numerischen Daten, 111
Einschränkungen für die Datenwerte, 112
Gefilterte Fälle, 117
mehrere Ansichten/Fenster, 118
mehrere geöffnete Datendateien, 121
Senden von Daten an andere Anwendungen, 687
Spaltenbreite, 107
Suchen von Fällen, 116–117
Variablenansicht, 97
Verschieben von Variablen, 116
Daten-Sets
umbenennen, 124
Datenanalyse, 14
grundlegende Schritte, 14
Datenansicht, 96
Datenbankdateien, 30–31, 35–37, 39–40, 43, 46, 48
Anmelden bei einer Datenbank, 35
Auswählen einer Datenquelle, 31
Auswählen von Datenfeldern, 36
bedingte Ausdrücke, 40

Datenbanksicherheit, 35
Definieren von Variablen, 46
Eigenschaften der Beziehung, 39
einlesen, 30–31, 36
Erstellen von Beziehungen, 37
Festlegen von Kriterien, 40
Microsoft Access, 31
Parameterabfragen, 40, 43
Speichern von Abfragen, 48
SQL-Syntax, 48
Überprüfen von Ergebnissen, 48
Umwandeln von Strings in numerische Variablen, 46
Verbindungen zwischen Tabellen, 37, 39
Wert abfragen, 43
Where-Klausel, 40
Zufallsstichproben, 40
Datendateien, 25–26, 50, 68–69, 74, 77, 81, 88, 90–91, 246
Dimensions, 62
Hinzufügen von Kommentaren, 658
Informationen aus dem Datenlexikon, 68–69
Informationen zur Datei, 68–69
Leistungssteigerung bei umfangreichen Dateien, 81
mehrere geöffnete Datendateien, 121
mrInterview, 62
öffnen, 25–26
Quancept, 62
Quanvert, 62
Remote-Server, 88, 90–91
schützen, 78
speichern, 69, 74
Speichern von Ausgaben als Datendateien im SPSS-Format, 747
Speichern von Untergruppen von Variablen, 77
Text, 50
transponieren, 223
umstrukturieren, 246
vertauschen, 223
Dateneingabe, 110
Datenlexikon, 68–69
Zuweisen aus einer anderen Datei, 138
Datentransformationen, 681
Bedingte Transformationen, 167
Berechnen von Variablen, 165
Bilden der Rangfolge, 181
Funktionen, 169
String-Variablen, 168
Umkodieren von Werten, 174–175, 177–178, 184
verzögerte Ausführung, 681
Zeitreihen, 206, 208
Datentypen, 101, 103, 116, 683
ändern, 116
Anzeigeformate, 103
definieren, 101
Eingabeformate, 103
spezielle Währung, 101, 683
Datumsformate, 101, 103, 681
zweistellige Jahresangaben, 681
Datumsvariablen
Addieren oder Subtrahieren zu bzw. von Datums-/Zeitvariablen, 188
Erstellen einer Datums-/Zeitvariablen aus einem String, 188
Erstellen einer Datums-/Zeitvariablen aus einem Variablen-Set, 188
Extrahieren eines Teils einer Datums-/Zeitvariablen, 188
für Zeitreihendaten definieren, 206

dBASE-Dateien, 25, 28, 74
öffnen, 25, 28
speichern, 74
Definieren von Variablen, 98, 101, 104, 106, 108–109, 126
Datentypen, 101
fehlende Werte, 106
Kopieren und Einfügen von Attributen, 108–109
Variablenlabels, 104
Vorlagen, 108–109
Wertelabels, 104, 126
Zuweisen eines Datenlexikons, 138
Dendrogramme
in der hierarchischen Clusteranalyse, 549
Deskriptive Statistiken, 371
Anzeigereihenfolge, 373
in der Explorativen Datenanalyse, 381
in der Two-Step-Clusteranalyse, 541
in Deskriptive Statistiken, 371
in GLM - Univariat, 458
in Häufigkeiten, 366
in Verhältnisstatistiken, 637
in Zusammenfassen, 403
Speichern von *Z*-Werten, 371
Statistiken, 373
zusätzliche Funktionen beim Befehl, 374
DfBeta
in Lineare Regression, 489
DfFit
in Lineare Regression, 489
Diagramme, 273, 281–282, 285, 673, 698
ändern, 643
Ausblenden, 273
Einfügen in andere Anwendungen, 282
erstellen, 639
exportieren, 285, 698
Fallbeschriftungen, 497
Fehlende Werte, 648
Fußnoten, 647
in ROC-Kurve, 653
Kopieren, 281
kopieren in andere Anwendungen, 281
Seitenverhältnis, 673
Titel, 647
Übersicht, 639
Untertitel, 647
Vorlagen, 651, 673
Diagramme mit der Streubreite gegen das mittlere Niveau
in der Explorativen Datenanalyse, 382
in GLM - Univariat, 458
Diagrammoptionen, 673
Dialogfelder, 12, 659, 661, 666, 735, 737
Anzeigen von Variablenlabels, 11, 666
Anzeigen von Variablennamen, 11, 666
Anzeigereihenfolge für Variablen, 666
Auswählen von Variablen, 12
Definieren von Variablen-Sets, 659
optionale Einstellungen, 12
Skripts, 735, 737
Steuerelemente, 11
Umsortieren von Listen der Zielvariablen, 662
untergeordnete Dialogfelder, 12
Variablen, 10
Variablenbeschreibung, 13
Variablensymbole, 13
Verwenden von Variablen-Sets, 661
Differenzen zwischen Gruppen
in OLAP-Würfel, 419
Differenzen zwischen Variablen
in OLAP-Würfel, 419
Differenzfunktion, 210

Differenzkontraste
in GLM, 450
Direkte Oblimin-Rotation
in der Faktorenanalyse, 527
Diskriminanzanalyse, 505
A-priori-Wahrscheinlichkeit, 512
Anzeigeoptionen, 511–512
Auswählen von Fällen, 508
Auswerten ausgeschlossener Fälle, 712
Beispiel, 505
Definieren eines Bereichs, 508
deskriptive Statistiken, 509
Diagramme, 512
Diskriminanzmethoden, 511
Exportieren von Modellinformationen, 514
fehlende Werte, 512
Funktionskoeffizienten, 509
Gruppenvariablen, 505
Kovarianzmatrix, 512
Kriterien, 511
Mahalanobis-Abstand, 511
Matrizen, 509
Rao-*V*, 511
schrittweise Methoden, 505
Speichern von Klassifikationsvariablen, 514
Statistiken, 505, 509
unabhängige Variablen, 505
Wilks-Lambda, 511
zusätzliche Funktionen beim Befehl, 515
Distanz nach Minkowski
in Distanzen, 477
Distanz nach Tschebyscheff
in Distanzen, 477
Distanzen, 475
Ähnlichkeitsmaße, 478
Beispiel, 475
Berechnen von Distanzen zwischen Fällen, 475
Berechnen von Distanzen zwischen Variablen, 475
Statistiken, 475
Transformieren von Maßen, 477–478
Transformieren von Werten, 477–478
Unähnlichkeitsmaße, 477
zusätzliche Funktionen beim Befehl, 479
Distanzmaße
in der hierarchischen Clusteranalyse, 547
in Distanzen, 477
Division
Dividieren über Berichtsspalten, 616
Dollarformat, 101, 103
Dollarzeichen, 343
in Pivot-Tabellen, 343
Doppelte Fälle (Datensätze)
finden und filtern, 149
Drehen von Beschriftungen, 321
Drucken, 120, 297, 299, 301–302, 304, 314–315, 333, 338, 350
Abstand zwischen Ausgabeobjekten, 304
Daten, 120
Diagramme, 297
Diagrammgröße, 304
Festlegen von Tabellenumbrüchen, 350
Kopf- und Fußzeilen, 301–302
Pivot-Tabellen, 297
Schichten, 297, 333, 338
Seitenansicht, 299
Seiteneinrichtung, 301
Seitennummern, 304
Skalieren von Tabellen, 333, 338
Textausgabe, 297, 314–315
Duncans multipler Spannweitentest
in einfaktorieller ANOVA, 437

in GLM, 453
Dunnett-*T*-Test
in einfaktorieller ANOVA, 437
in GLM, 453
Durbin-Watson-Statistik
in Lineare Regression, 493
Durchschnittliche absolute Abweichung (AAD)
in Verhältnisstatistiken, 637

Ehrlich signifikante Differenz nach Tukey
in einfaktorieller ANOVA, 437
in GLM, 453
Eigenschaften, 332–333, 730
OLE-Automatisierungsobjekte, 730
Pivot-Tabellen, 332
Tabellen, 333
Eigenwerte
in der Faktorenanalyse, 523–524
in Lineare Regression, 493
Einbetten
Interaktive Diagramme, 282
Pivot-Tabellen, 282
Einblenden, 273, 328–329, 689
Dimensionsbeschriftungen, 329
Ergebnisse, 273
Erklärungen, 329
Fußnoten, 329
Symbolleisten, 689
Titel, 329
Zeilen oder Spalten, 328
Einfache Kontraste
in GLM, 450
Einfaktorielle ANOVA, 433
Faktorvariablen, 433
fehlende Werte, 440
Kontraste, 436
Mehrfachvergleiche, 437
Optionen, 440
Polynomiale Kontraste, 436
Post-Hoc-Tests, 437
Statistiken, 440
zusätzliche Funktionen beim Befehl, 441
einfügen, 282–284
Diagramme, 282
Pivot-Tabellen, 282–283
Pivot-Tabellen als Tabellen, 283
spezielle Objekte, 284
Einfügen (Schaltfläche), 11
Einfügen von Gruppenbeschriftungen, 320
Eingabeformate, 103
Eingeben von Daten, 110–112
nichtnumerisch, 111
numerisch, 111
Verwenden von Wertelabels, 112
Eiszapfendiagramme
in der hierarchischen Clusteranalyse, 549
Entfernen von Gruppenbeschriftungen, 321
EPS-Dateien, 285, 292, 295, 698
Exportieren von Diagrammen, 285, 292, 295, 698
Equamax-Rotation
in der Faktorenanalyse, 527
Erklärungen, 349
Hinzufügen zu einer Tabelle, 349
Ersetzen fehlender Werte
lineare Interpolation, 213
linearer Trend, 213
Median der Nachbarpunkte, 213
Mittel der Nachbarpunkte, 213
Mittelwert der Datenreihe, 213
Erste
in Mittelwerte, 410

in OLAP-Würfel, 416
in Zusammenfassen, 403
Erwartete Anzahl
in Kreuztabellen, 395
Eta
in Kreuztabellen, 391
in Mittelwerte, 410
Eta-Quadrat
in GLM - Univariat, 458
in Mittelwerte, 410
Euklidische Distanz
in Distanzen, 477
Exakter Test nach Fisher
in Kreuztabellen, 391
Excel-Dateien, 25, 27, 74, 687
Hinzufügen eines Menüeintrags zum Senden von Daten an Excel, 687
öffnen, 25, 27
speichern, 74
Excel-Format
Exportieren von Ausgaben, 285, 289
EXECUTE-Befehl
Einfügen aus Dialogfeldern, 362
Explorative Datenanalyse, 377
Diagramme, 382
fehlende Werte, 384
Optionen, 384
Potenztransformationen, 383
Statistiken, 381
zusätzliche Funktionen beim Befehl, 384
Exponentielles Modell
in Kurvenanpassung, 501
Exportieren von Ausgaben, 285, 289–290, 698, 710
Excel-Format, 285, 289
HTML, 289
OMS, 747
PowerPoint-Format, 285
Publizieren im Internet, 710
Word-Format, 285, 289
Exportieren von Daten, 687
Hinzufügen eines Menüeintrags zum Exportieren von Daten, 687
Exportieren von Diagrammen, 285, 292–295, 297, 695, 698
automatisierte Produktion, 695
Diagrammgröße, 292
Extremwerte
in der Explorativen Datenanalyse, 381

F nach R-E-G-W
in einfaktorieller ANOVA, 437
in GLM, 453
Faktorenanalyse, 517
Anzeigeformat für Koeffizienten, 529
Auswählen von Fällen, 523
Auswerten ausgeschlossener Fälle, 712
Beispiel, 517
deskriptive Statistiken, 523
Extraktionsmethoden, 524
Faktorwerte, 528
fehlende Werte, 529
Konvergenz, 524, 527
Ladungsdiagramme, 527
Rotationsmethoden, 527
Statistiken, 517, 523
Übersicht, 517
zusätzliche Funktionen beim Befehl, 530
Faktorwerte, 528
Fälle, 115–117, 246
Auffinden doppelt vorhandener, 149
Auswählen von Teilmengen, 238, 240, 242

Einfügen von neuen Fällen, 115
gewichten, 243
sortieren, 222
Suchen im Daten-Editor, 116–117
Umstrukturieren in Variablen, 246
Fallweise Diagnose
in Lineare Regression, 493
Farben in Pivot-Tabellen, 337, 339, 346
Rahmen, 337
Schriftart, 339
Zellenhintergrund, 346
Zellenvordergrund, 346
Fehlende Werte, 106
definieren, 106
im Sequenzentest, 572
in Bericht in Zeilen, 610
in Berichte in Spalten, 618
in bivariaten Korrelationen, 466
in Chi-Quadrat-Test, 565
in der Explorativen Datenanalyse, 384
in der Faktorenanalyse, 529
in Diagrammen, 648
in einfaktorieller ANOVA, 440
in Funktionen, 170
in Kolmogorov-Smirnov-Test bei einer Stichprobe, 575
in Lineare Regression, 495
in Mehrfachantworten: Häufigkeiten, 596
in Mehrfachantworten: Kreuztabellen, 601
in Partielle Korrelationen, 472
in ROC-Kurve, 656
in T-Test bei einer Stichprobe, 430
in T-Test bei gepaarten Stichproben, 428
in T-Test bei unabhängigen Stichproben, 425
in Test auf Binomialverteilung, 568
in Tests bei mehreren unabhängigen Stichproben, 587
in Tests bei zwei unabhängigen Stichproben, 579
in Tests bei zwei verbundenen Stichproben, 583
in Zeitreihendaten ersetzen, 211
String-Variablen, 106
Fehlersuche in Skripts, 740, 742
Abarbeiten von Skripts, 740
Debug-Fenster, 742
Haltepunkte, 740
Fenster, 7
Aktives Fenster, 8
Hauptfenster, 8
Fensterteiler
Daten-Editor, 118
Festes Format, 50
Festlegen der Größe von exportierten Diagrammen, 292
Formatierung, 309
Spalten in Berichten, 607
Textausgabe, 309
Fortsetzungstext, 338
für Pivot-Tabellen, 338
Freies Format, 50
Freigegebene Laufwerke, 93
Friedman-Test
in Tests bei mehreren verbundenen Stichproben, 590
Funktion für gleitenden Median, 210
Funktion für kumulierte Summe, 210
Funktion für saisonale Differenz, 210
Funktion für zentrierten gleitenden Durchschnitt, 210
Funktion für zurückgreifenden gleitenden Durchschnitt, 210

Funktionen, 169
  Behandlung fehlender Werte, 170
Fußnoten, 334, 346–347, 349
  Hinzufügen zu einer Tabelle, 349
  in Diagrammen, 647
  Markierungen, 334, 346
  neu numerieren, 347
Fußzeilen, 301–302

Gamma
  in Kreuztabellen, 391
Gefilterte Fälle, 117
  im Daten-Editor, 117
Geometrisches Mittel
  in Mittelwerte, 410
  in OLAP-Würfel, 416
  in Zusammenfassen, 403
Geringste signifikante Differenz
  in einfaktorieller ANOVA, 437
  in GLM, 453
Gesamtergebnisse
  in Berichte in Spalten, 618
Gesamtergebnisspalte
  in Berichten, 616
Gesamtprozentwerte
  in Kreuztabellen, 395
Gesamtsummen, 712
  automatischer Fettdruck in der Ausgabe, 712
Gesättigte Modelle
  in GLM, 447
Geschätzte Randmittel
  in GLM - Univariat, 458
Geschwindigkeit, 81
  Zwischenspeichern von Daten, 81
Get Data (Befehlssyntax), 78
  Vergleich mit 'Data List', 78
  Vergleich mit 'Get Capture', 78
Getrimmtes Mittel
  in der Explorativen Datenanalyse, 381
Gewichten von Fällen, 243
  nicht-ganzzahlige Gewichtungen in Kreuztabellen, 243
Gewichtete Daten, 269
  und umstrukturierte Datendateien, 269
Gewichtete kleinste Quadrate
  in Lineare Regression, 481
Gewichtete Schätzwerte
  in GLM, 456
Gewichteter Mittelwert
  in Verhältnisstatistiken, 637
Gitterlinien, 338
  Pivot-Tabellen, 338
Glättungsfunktion, 210
Gliederung, 276, 278
  Ändern von Ebenen, 278
  erweitern, 278
  im Viewer, 276
  reduzieren, 278
GLM
  Modell, 447
  Post-Hoc-Tests, 453
  Profilplots, 451
  Quadratsumme, 447
  Speichern von Matrizen, 456
  Speichern von Variablen, 456
GLM - Univariat, 443, 460
  anzeigen, 458
  Diagnose, 458
  geschätzte Randmittel, 458
  Kontraste, 450
  Optionen, 458
Globale Prozeduren, 684, 734

Globale Skripts, 734
Goodman-und-Kruskal-Gamma
in Kreuztabellen, 391
Goodman-und-Kruskal-Lambda
in Kreuztabellen, 391
Goodman-und-Kruskal-Tau
in Kreuztabellen, 391
Größen, 279
in der Gliederung, 279
Größendifferenzmaß
in Distanzen, 477
Gruppenbeschriftungen, 320
Gruppenmittelwerte, 407, 413
Gruppenvariablen, 246
erstellen, 246
Gruppieren von Zeilen oder Spalten, 320
Gruppierter Median
in Mittelwerte, 410
in OLAP-Würfel, 416
in Zusammenfassen, 403
GT2 nach Hochberg
in einfaktorieller ANOVA, 437
in GLM, 453
Guttman-Modelle
in der Reliabilitätsanalyse, 621, 623

Haltepunkte, 740
in Skripts, 740
Harmonisches Mittel
in Mittelwerte, 410
in OLAP-Würfel, 416
in Zusammenfassen, 403
Häufigkeiten, 363
Anzeigereihenfolge, 369
Diagramme, 368
Formate, 369
Statistiken, 366
Unterdrücken von Tabellen, 369
Häufigkeitstabellen
in der Explorativen Datenanalyse, 381
in Häufigkeiten, 363
Hauptachsen-Faktorenanalyse, 524
Hauptfenster, 8
Hauptkomponentenanalyse, 517, 524
Hebelwerte
in GLM, 456
in Lineare Regression, 489
Helmert-Kontraste
in GLM, 450
Hierarchische Clusteranalyse, 543
Ähnlichkeitsmaße, 547
Beispiel, 543
Cluster-Methoden, 547
Cluster-Zugehörigkeit, 548–549
Clustern von Fällen, 543
Clustern von Variablen, 543
Dendrogramme, 549
Diagrammausrichtung, 549
Distanzmaße, 547
Distanzmatrizen, 548
Eiszapfendiagramme, 549
Speichern von neuen Variablen, 549
Statistiken, 543, 548
Transformieren von Maßen, 547
Transformieren von Werten, 547
Zuordnungsübersichten, 548
zusätzliche Funktionen beim Befehl, 550
Hierarchische Zerlegung, 448
Hilfe (Schaltfläche), 11
Hilfe-Fenster, 17
Hinzufügen von Gruppenbeschriftungen, 320

Histogramme
in der Explorativen Datenanalyse, 382
in Häufigkeiten, 368
in Lineare Regression, 487
Höchstzahl Verzweigungen
in der Two-Step-Clusteranalyse, 537
Hotellings T^2
in der Reliabilitätsanalyse, 621, 623
HTML, 285, 289, 698, 783
Exportieren von Ausgaben, 285, 289, 698
Hinzufügen von benutzerdefiniertem Code, 783

ICC. *Siehe* Korrelationskoeffizienten in Klassen, 623
Image-Faktorisierung, 524
Immediate (Registerkarte), 742
Skript-Fenster, 742
in Zwischenspeicher ablegen, 81
Aktive Datei, 81
Informationen zur Datei, 68–69
Interaktive Diagramme, 281–282, 678
Einbetten als ActiveX-Objekte, 282
kopieren in andere Anwendungen, 281
Optionen, 678
Speichern von Daten mit den Diagrammen, 678
Internet, 710
Publizieren von Ausgaben, 710
Inverses Modell
in Kurvenanpassung, 501
Iterationen
in der Clusterzentrenanalyse, 557
in der Faktorenanalyse, 524, 527

Jahre, 681
zweistellige Angaben, 681
Journaldatei, 666

JPEG-Dateien, 285, 292, 698
Exportieren von Diagrammen, 285, 292, 698

Kappa
in Kreuztabellen, 391
Kategoriale Daten, 131
Umwandeln von Intervalldaten in diskrete Kategorien, 152
Kendall-Tau-*b*
in bivariaten Korrelationen, 463
in Kreuztabellen, 391
Kendall-Tau-*c*, 391
in Kreuztabellen, 391
Kendall-*W*
in Tests bei mehreren verbundenen Stichproben, 590
Klassifikation
in ROC-Kurve, 653
Kollinearitätsdiagnose
in Lineare Regression, 493
Kolmogorov-Smirnov-Test bei einer Stichprobe, 573
Fehlende Werte, 575
Optionen, 575
Statistiken, 575
zu testende Verteilung, 573
zusätzliche Funktionen beim Befehl, 575
Kolmogorov-Smirnov-*Z*
in Kolmogorov-Smirnov-Test bei einer Stichprobe, 573
in Tests bei zwei unabhängigen Stichproben, 578
Komma-getrennte Daten, 50
Kommaformat, 101, 103
Konfidenzintervalle
in der Explorativen Datenanalyse, 381
in einfaktorieller ANOVA, 440

in GLM, 450, 458
in Lineare Regression, 493
in ROC-Kurve, 656
in T-Test bei einer Stichprobe, 430
in T-Test bei gepaarten Stichproben, 428
in T-Test bei unabhängigen Stichproben, 425
Speichern in Lineare Regression, 489
Konstruieren von Termen, 448
Kontextsensitive Hilfe, 322
Suchen von Definitionen für Beschriftungen in Pivot-Tabellen, 322
Kontingenzkoeffizient
in Kreuztabellen, 391
Kontingenztafeln, 387
Kontinuitätskorrektur nach Yates
in Kreuztabellen, 391
Kontraste
in einfaktorieller ANOVA, 436
in GLM, 450
Kontroll-Variablen
in Kreuztabellen, 390
Konvergenz
in der Clusterzentrenanalyse, 557
in der Faktorenanalyse, 524, 527
Konzentrationsindex
in Verhältnisstatistiken, 637
Kopfzeilen, 301–302
Kopieren, 275, 281
Ausgabe, 275
Diagramme, 281
Pivot-Tabellen, 281
Korrelationen
in bivariaten Korrelationen, 463
in Kreuztabellen, 391
in Partielle Korrelationen, 469
nullter Ordnung, 472
Korrelationen nullter Ordnung
in Partielle Korrelationen, 472
Korrelationskoeffizient nach Spearman
in bivariaten Korrelationen, 463
in Kreuztabellen, 391
Korrelationskoeffizienten in Klassen (ICC)
in der Reliabilitätsanalyse, 623
Korrelationsmatrix
in der Diskriminanzanalyse, 509
in der Faktorenanalyse, 517, 523
korrigiertes R^2
in Lineare Regression, 493
Kovarianzmatrix
in der Diskriminanzanalyse, 509, 512
in GLM, 456
in Lineare Regression, 493
Kovarianzverhältnis
in Lineare Regression, 489
KR20
in der Reliabilitätsanalyse, 623
Kreisdiagramme
in Häufigkeiten, 368
Titel, 647
Kreuztabelle
in Kreuztabellen, 387
Mehrfachantworten, 598
Kreuztabellen, 387
Formate, 397
Gruppiertes Balkendiagramm, 391
Kontroll-Variablen, 390
nicht-ganzzahlige Gewichtungen, 243
Schichten, 390
Statistiken, 391
Unterdrücken von Tabellen, 387
Zellen anzeigen, 395

Kruskal-Tau
in Kreuztabellen, 391
Kruskal-Wallis-*H*
in Tests bei zwei unabhängigen Stichproben, 584
Kubisches Modell
in Kurvenanpassung, 501
Kuder-Richardson-20 (KR20)
in der Reliabilitätsanalyse, 623
Kurtosis
in Bericht in Spalten, 615
in Bericht in Zeilen, 608
in der Explorativen Datenanalyse, 381
in Deskriptive Statistiken, 373
in Häufigkeiten, 366
in Mittelwerte, 410
in OLAP-Würfel, 416
in Zusammenfassen, 403
Kurvenanpassung, 497
Einschließen von Konstanten, 497
Modelle, 501
Prognose, 502
Speichern von Residuen, 502
Speichern von Vorhersageintervallen, 502
Speichern vorhergesagter Werte, 502
Varianzanalyse, 497

Ladungsdiagramme
in der Faktorenanalyse, 527
LAG (Funktion), 210
Lagemaße
in der Explorativen Datenanalyse, 381
in Häufigkeiten, 366
in Verhältnisstatistiken, 637
Lambda
in Kreuztabellen, 391
Leerzeichen-getrennte Daten, 50
Leistung, 81
Zwischenspeichern von Daten, 81
Lesezeichen, 326
Letzte
in Mittelwerte, 410
in OLAP-Würfel, 416
in Zusammenfassen, 403
Levene-Test
in der Explorativen Datenanalyse, 382
in einfaktorieller ANOVA, 440
in GLM - Univariat, 458
Likelihood-Quotienten-Chi-Quadrat
in Kreuztabellen, 391
Lilliefors-Test
in der Explorativen Datenanalyse, 382
Lineare Regression, 481
Auswahlmethoden für Variablen, 485, 495
Auswahlvariable, 487
Blöcke, 481
Diagramme, 487
Exportieren von Modellinformationen, 489
Fehlende Werte, 495
Gewichtungen, 481
Residuen, 489
Speichern von neuen Variablen, 489
Statistiken, 493
zusätzliche Funktionen beim Befehl, 496
Lineares Modell
in Kurvenanpassung, 501
Linearitätstests
in Mittelwerte, 410
Liste der Zielvariablen, 662
Loaded (Registerkarte), 742
Skript-Fenster, 742
Logarithmisches Modell
in Kurvenanpassung, 501

Logistisches Modell
in Kurvenanpassung, 501
Löschen mehrerer EXECUTE-Befehle in Syntaxdateien, 362
Löschen von Ausgaben, 274
Lotus 1-2-3-Dateien, 25, 74, 687
Hinzufügen eines Menüeintrags zum Senden von Daten an Lotus, 687
öffnen, 25
speichern, 74
LSD nach Fisher
in GLM, 453

M-Schätzer
in der Explorativen Datenanalyse, 381
M-Schätzer nach Hampel
in der Explorativen Datenanalyse, 381
M-Schätzer nach Huber
in der Explorativen Datenanalyse, 381
Mahalanobis-Abstand
in der Diskriminanzanalyse, 511
in Lineare Regression, 489
Mann-Whitney-*U*-Test
in Tests bei zwei unabhängigen Stichproben, 578
Mantel-Haenszel-Statistik
in Kreuztabellen, 391
Maßeinheit, 666
Maximum
in der Explorativen Datenanalyse, 381
in Deskriptive Statistiken, 373
in Häufigkeiten, 366
in Mittelwerte, 410
in OLAP-Würfel, 416
in Verhältnisstatistiken, 637
in Zusammenfassen, 403
Vergleichen von Berichtsspalten, 616
Maximum Likelihood
in der Faktorenanalyse, 524
McNemar-Test
in Kreuztabellen, 391
in Tests bei zwei verbundenen Stichproben, 580
Median
in der Explorativen Datenanalyse, 381
in Häufigkeiten, 366
in Mittelwerte, 410
in OLAP-Würfel, 416
in Verhältnisstatistiken, 637
in Zusammenfassen, 403
Mediantest
in Tests bei zwei unabhängigen Stichproben, 584
mehrere Ansichten/Fenster
Daten-Editor, 118
Mehrere geöffnete Datendateien, 121
Mehrfachantworten
zusätzliche Funktionen beim Befehl, 603
Mehrfachantworten: Häufigkeiten, 596
Fehlende Werte, 596
Mehrfachantworten: Kreuztabellen, 598
Definieren von Wertebereichen, 601
Fehlende Werte, 601
Paaren von Variablen aus verschiedenen Antworten-Sets, 601
Prozentwerte basierend auf Antworten, 601
Prozentwerte basierend auf Fällen, 601
Prozentwerte für Zellen, 601
Mehrfachantworten-Sets
definieren, 134
mehrere Kategorien, 134
Set aus dichotomen Variablen, 134
Mehrfachantworten-Sets definieren, 594
Dichotomien, 594
Kategorien, 594

Set-Labels, 594
Set-Namen, 594
Mehrfache Regression
in Lineare Regression, 481
Mehrfachvergleiche
in einfaktorieller ANOVA, 437
Menüs, 9, 687
Anpassen, 687
Meßniveau, 99, 131
definieren, 99
Symbole in Dialogfeldern, 13
Metadateien, 285, 292, 698
Exportieren von Diagrammen, 285, 292, 698
Methoden, 730
OLE-Automatisierungsobjekte, 730
Metrisch, 99
Meßniveau, 99
Metrische Variablen
in Bereiche einteilen, um kategoriale Variablen zu erstellen, 152
Microsoft Access, 31
Minimum
in der Explorativen Datenanalyse, 381
in Deskriptive Statistiken, 373
in Häufigkeiten, 366
in Mittelwerte, 410
in OLAP-Würfel, 416
in Verhältnisstatistiken, 637
in Zusammenfassen, 403
Vergleichen von Berichtsspalten, 616
Mittelwert
in Bericht in Spalten, 615
in Bericht in Zeilen, 608
in der Explorativen Datenanalyse, 381
in Deskriptive Statistiken, 373
in einfaktorieller ANOVA, 440
in Häufigkeiten, 366
in Mittelwerte, 410
in OLAP-Würfel, 416
in Verhältnisstatistiken, 637
in Zusammenfassen, 403
Untergruppe, 407, 413
von mehreren Berichtsspalten, 616
Mittelwerte, 407, 712
ASCII-Ausgabe aus, 712
Optionen, 410
Statistiken, 410
Mittelwerte von Untergruppen, 407, 413
Modalwert
in Häufigkeiten, 366
Modelldatei
Laden von gespeicherten Modellen zum Bewerten von Daten, 215
mrInterview, 62
Multidimensionale Skalierung , 627
Anzeigeoptionen, 633
Beispiel, 627
Definieren der Datenform, 630
Dimensionen, 631
Distanzmaße, 630
Erstellen von Distanzmatrizen, 630
Konditionalität, 631
Kriterien, 633
Meßniveaus, 631
Skalierungsmodelle, 631
Statistiken, 627
Transformieren von Werten, 630
zusätzliche Funktionen beim Befehl, 633
Multipler Spannweitentest nach Ryan-Einot-Gabriel-Welsch
in einfaktorieller ANOVA, 437
in GLM, 453

Multiples *F* nach Ryan-Einot-Gabriel-Welsch
in einfaktorieller ANOVA, 437
in GLM, 453
Multiples *R*
in Lineare Regression, 493
Multiplikation
Multiplizieren über Berichtsspalten, 616
Muster-Differenzmaß
in Distanzen, 477
Mustermatrix
in der Faktorenanalyse, 517

Neue Funktionen
SPSS 14.0, 2
Newman-Keuls
in GLM, 453
Nicht standardisierte Residuen
in GLM, 456
Nichtparametrische Tests
Chi-Quadrat, 562
Kolmogorov-Smirnov-Test bei einer Stichprobe, 573
Sequenzentest, 569
Tests bei mehreren unabhängigen Stichproben, 584
Tests bei mehreren verbundenen Stichproben, 588
Tests bei zwei unabhängigen Stichproben, 576
Tests bei zwei verbundenen Stichproben, 580
Nominal, 99
Meßniveau, 99, 131
Normalrangwerte
beim Bilden der Rangfolge, 182
Normalverteilungsdiagramme
in der Explorativen Datenanalyse, 382
in Lineare Regression, 487
Numerisches Format, 101, 103

Objekte, 725, 727–728, 731
Konventionen für die Benennung von Variablen, 727
Übersicht, 725, 727
verwenden in Skripts, 725, 728, 731
Objektkatalog, 731
Öffnen von Dateien, 25–31, 50
Datendateien, 25–26
dBASE-Dateien, 25, 28
Excel-Dateien, 25, 27
Lotus 1-2-3-Dateien, 25
Stata-Dateien, 29
SYSTAT-Dateien, 25
Tabellenkalkulationsdateien, 25, 28
Tabulator-getrennte Dateien, 25
Textdatendateien, 50
OK (Schaltfläche), 11
OLAP-Würfel, 413
Statistiken, 416
Titel, 420
OLE-Automatisierung, 711, 725, 727–728, 730–731
Eigenschaften, 730
Erstellen von Skripts mit, 711
Konventionen für die Benennung von Variablen, 727
Methoden, 730
Übersicht, 725, 727–728, 731
Verwenden von Objekten, 725, 728, 731
OMS, 747, 778
Ausgabe aus dem Viewer ausschließen, 763
Ausgabeobjekttypen, 751
Befehls-IDs, 753
SAV-Dateiformat, 757, 763

SPSS-Datendateiformat, 757, 763
Steuern von Pivots in Tabellen, 757, 770
Tabellenuntertypen, 753
Textformat, 757
Variablennamen in SAV-Dateien, 772
Verwenden von XSLT mit OXML, 779
XML, 757, 774
Online-Hilfe, 17
Statistik-Assistent, 15
Optionen, 666, 668–669, 671, 673, 678–679, 681, 683–684
allgemein, 666
Beschriftung der Ausgabe, 671
Daten, 681
Diagramme, 673
Interaktive Diagramme, 678
Skripts, 684
Text-Viewer, 669
Verzeichnis für temporäre Dateien, 666
Viewer, 668
Vorlage für Pivot-Tabellen, 679
Währung, 683
zweistellige Jahresangaben, 681
Ordinal, 99
Meßniveau, 99, 131
OXML, 779

Paarweiser Vergleichstest nach Gabriel
in einfaktorieller ANOVA, 437
in GLM, 453
Paarweiser Vergleichstest nach Games und Howell
in einfaktorieller ANOVA, 437
in GLM, 453
Paralleles Modell
in der Reliabilitätsanalyse, 621, 623
Parameterschätzer
in GLM - Univariat, 458
Partielle Diagramme
in Lineare Regression, 487
Partielle Korrelationen, 469
Fehlende Werte, 472
in Lineare Regression, 493
Korrelationen nullter Ordnung, 472
Optionen, 472
Statistiken, 472
zusätzliche Funktionen beim Befehl, 473
Paßwortschutz, 305
Pearson-Chi-Quadrat
in Kreuztabellen, 391
Pearson-Korrelation
in bivariaten Korrelationen, 463
in Kreuztabellen, 391
Perzentile
in der Explorativen Datenanalyse, 381
in Häufigkeiten, 366
Phi
in Kreuztabellen, 391
Phi-Quadrat-Distanzmaß
in Distanzen, 477
PICT-Dateien, 285, 292–293, 698
Exportieren von Diagrammen, 285, 292–293, 698
Pivot-Tabellen, 273, 280–285, 297, 317–322, 327, 330, 332–335, 337–338, 340, 348–350, 679, 698, 704
allgemeine Eigenschaften, 333
Ändern der Anzeigereihenfolge, 319–320
Ändern der Darstellung, 330
Ändern der Vorlage, 330
Anpassen an die Seitengröße, 333, 338
Anpassen der Standard-Spaltenbreite, 679

Anzeigen ausgeblendeter Rahmen, 338
Aufheben der Gruppierung von Zeilen oder Spalten, 321
Ausblenden, 273
Auswählen von Zeilen und Spalten, 348
bearbeiten, 317
Bearbeiten von zwei oder mehr, 317
Drehen von Beschriftungen, 321
Drucken umfangreicher Tabellen, 350
Drucken von Schichten, 297
Eigenschaften, 332
Ein- und Ausblenden von Zellen, 327
Einbetten als ActiveX-Objekte, 282
Einfügen als Metadateien, 282
Einfügen als Tabellen, 280, 283
Einfügen als Text, 283
Einfügen in andere Anwendungen, 280
Einfügen von Gruppenbeschriftungen, 320
exportieren als HTML, 285, 698
Festlegen von Formaten für Produktionsjobs, 704
Festlegen von Tabellenumbrüchen, 350
Fortsetzungstext, 338
Fußnoteneigenschaften, 334
Gitterlinien, 338
Gruppieren von Zeilen oder Spalten, 320
Hinzufügen von Erklärungen, 349
Identifizieren von Dimensionen, 319
Kopieren, 281
kopieren in andere Anwendungen, 281
Kopieren und Einfügen mehrerer Tabellen, 284
Löschen von Gruppenbeschriftungen, 321
pivotieren, 317–318
Rahmen, 337
Schichten, 322
Standardvorlage für neue Tabellen, 679
Suchen von Definitionen für Beschriftungen, 322
Transponieren von Zeilen und Spalten, 319
Verschieben von Zeilen und Spalten, 320
Verwenden von Symbolen, 318
Zellenbreiten, 340
Zellenformate, 335
Zurücksetzen auf die Standardeinstellungen, 322
pivotieren
mit OMS für exportierte Ausgabe steuern, 770
PNG-Dateien, 285, 294
Exportieren von Diagrammen, 285, 294
Polynomiale Kontraste
in einfaktorieller ANOVA, 436
in GLM, 450
portable Dateien
Variablennamen, 74
Portnummern, 86
Post-Hoc-Mehrfachvergleiche, 437
PostScript-Dateien (Encapsulated), 285, 295, 698
Exportieren von Diagrammen, 285, 295, 698
Potenzmodell
in Kurvenanpassung, 501
PowerPoint, 289
Exportieren von Ausgaben als PowerPoint, 289
PowerPoint-Format
Exportieren von Ausgaben, 285
Preisbezogenes Differential (PRD)
in Verhältnisstatistiken, 637
Profilplots
in GLM, 451
Prognose
in Kurvenanpassung, 502
Programmieren mit Befehlssprache, 353
Prozeduren, 732
Skripts, 732

Prozeduren in Funktionen, 732
Prozeduren in Subroutinen, 732
Prozentwerte
in Kreuztabellen, 395
Prozentzeichen, 343
in Pivot-Tabellen, 343
Publizieren von Ausgaben, 710
mit dem SPSS-Produktionsmodus, 709
Punktformat, 101, 103

Q nach R-E-G-W
in einfaktorieller ANOVA, 437
in GLM, 453
Quadratisches Modell
in Kurvenanpassung, 501
Quadratsumme, 448
in GLM, 447
Quadrierte Euklidische Distanz
in Distanzen, 477
Quancept, 62
Quanvert, 62
Quartile
in Häufigkeiten, 366
Quartimax-Rotation
in der Faktorenanalyse, 527

r-Korrelationskoeffizient
in bivariaten Korrelationen, 463
in Kreuztabellen, 391
R-Statistik
in Lineare Regression, 493
in Mittelwerte, 410
R^2
Änderung in R^2, 493
in Lineare Regression, 493
in Mittelwerte, 410
Rahmen, 309, 337–338
Anzeigen ausgeblendeter Rahmen, 338
Text-Viewer, 309
Ränder, 301, 345
in Zellen, 345
Rang-Korrelationskoeffizient
in bivariaten Korrelationen, 463
Rankit-Schätzungen, 182
Rao-*V*
in der Diskriminanzanalyse, 511
Rauschverarbeitung
in der Two-Step-Clusteranalyse, 537
Reduzieren von Kategorien, 152
Referenzkategorie
in GLM, 450
Regression
Diagramme, 487
Lineare Regression, 481
Mehrfache Regression, 481
Regressionskoeffizienten
in Lineare Regression, 493
Relatives Risiko
in Kreuztabellen, 391
Reliabilitätsanalyse, 621
ANOVA-Tabelle, 623
Beispiel, 621
deskriptive Statistiken, 623
Hotellings T^2, 623
Inter-Item-Korrelationen und -Kovarianzen, 623
Korrelationskoeffizienten in Klassen, 623
Kuder-Richardson-20, 623
Statistiken, 621, 623
Tukeys Additivitätstest, 623
zusätzliche Funktionen beim Befehl, 626
Remote-Server, 83–86, 88, 90–91, 93, 703
Anmelden beim, 85

bearbeiten, 86
hinzufügen, 86
Speichern von Datendateien, 90
SPSS-Produktionsmodus, 703
UNC-Pfade, 93
verfügbare Prozeduren, 93
Zugriff auf Datendateien, 88, 91
Residuen
in Kreuztabellen, 395
Speichern in Kurvenanpassung, 502
Speichern in Lineare Regression, 489
Residuen-Diagramme
in GLM - Univariat, 458
Rho
in bivariaten Korrelationen, 463
in Kreuztabellen, 391
Risiko
in Kreuztabellen, 391
ROC-Kurve, 653
Statistiken und Diagramme, 656
Rückwärtselimination
in Lineare Regression, 485

S-Modell
in Kurvenanpassung, 501
S-Streß
in Multidimensionale Skalierung, 627
SAS-Dateien
öffnen, 25
speichern, 74
SAV-Dateiformat
Ausgabe an eine SPSS-Datendatei weiterleiten, 757, 763
Savage-Werte, 182
Schaltflächen, 693
Bearbeiten von Bitmap-Symbolen für Symbolleisten, 693
Schattierung, 346
in Zellen, 346
Schätzer der Effektgröße
in GLM - Univariat, 458
Schätzer der Schärfe
in GLM - Univariat, 458
Scheffé-Test
in einfaktorieller ANOVA, 437
in GLM, 453
Schichten, 297, 322, 324, 333, 338
ändern, 324
anzeigen, 322, 324
Drucken, 297, 333, 338
erstellen, 322
in Kreuztabellen, 390
in Pivot-Tabellen, 322
Schiefe
in Bericht in Spalten, 615
in Bericht in Zeilen, 608
in der Explorativen Datenanalyse, 381
in Deskriptive Statistiken, 373
in Häufigkeiten, 366
in Mittelwerte, 410
in OLAP-Würfel, 416
in Zusammenfassen, 403
Schriftarten, 118, 279, 314, 339
Farben, 339
im Daten-Editor, 118
im Gliederungsfenster, 279
in Text-Viewer, 314
in Zellen, 339
Schrittweise Auswahl
in Lineare Regression, 485

Seiteneinrichtung, 301–302, 304
  Diagrammgröße, 304
  Kopf- und Fußzeilen, 302
Seiteneinstellung
  in Berichte in Spalten, 617
  in Berichte in Zeilen, 610
Seitennumerierung, 301, 304
  in Berichte in Spalten, 618
  in Berichte in Zeilen, 610
Seitenränder, 301
Seitenverhältnis, 673
Sequenzentest, 569
  Fehlende Werte, 572
  Optionen, 572
  Statistiken, 572
  Trennwerte, 569, 571
  zusätzliche Funktionen beim Befehl, 572
Server, 85–86
  Anmelden beim, 85
  bearbeiten, 86
  hinzufügen, 86
  Namen, 86
  Portnummern, 86
Shapiro-Wilk-Test
  in der Explorativen Datenanalyse, 382
Sidak-*T*-Test
  in einfaktorieller ANOVA, 437
  in GLM, 453
Sitzungs-Journal, 666
Skalieren
  in der Reliabilitätsanalyse, 621
  in Multidimensionale Skalierung, 627
  Meßniveau, 131
Skalieren von exportierten Diagrammen, 292
Skalieren von Pivot-Tabellen, 333, 338
Skript-Fenster, 716, 718, 731
  Eigenschaften, 718
  Immediate (Registerkarte), 742
  Loaded (Registerkarte), 742
  Menü “Debug”, 740
  Objektkatalog, 731
  Stack (Registerkarte), 742
  Watch (Registerkarte), 742
Skripts, 684, 687, 692, 711–713, 715, 718, 720, 735, 740
  ausführen, 711
  Ausführen mit Symbolleisten-Schaltflächen, 692
  Autoskript-Datei, 684, 724
  Autoskripts, 713, 720, 724
  Datei für globale Prozeduren, 684, 734
  Deklarieren von Variablen, 726–727
  Dialogfelder, 735, 737
  erstellen, 715, 720
  Fehlersuche, 740, 742
  Hinzufügen einer Beschreibung, 735
  Hinzufügen zu Menüs, 687
  mit Befehlssyntax, 743, 745
  Skript-Fenster, 716, 718
  Starter-Skripts, 718
  Übersicht, 711
  Verwenden von Automatisierungsobjekten, 725, 727–728, 731
Somers-*d*
  in Kreuztabellen, 391
Sortieren von Fällen, 222
Spalten, 341, 348
  Ändern der Breite in Pivot-Tabellen, 341
  Auswählen in Pivot-Tabellen, 348
Spaltenbreite, 107, 333, 341, 679
  Beeinflussen der Breite bei Texten mit Zeilenumbrüchen, 333

Einstellen der Standardbreite, 679
Festlegen der maximalen Breite, 333
im Daten-Editor, 107
Pivot-Tabellen, 341
Spaltenformat, 343
Ändern in Pivot-Tabellen, 343
Spaltenprozente
in Kreuztabellen, 395
Spearman-Brown-Reliabilität
in der Reliabilitätsanalyse, 623
Speichern der Ausgabe, 285, 289–290, 305, 316, 698, 710
Excel-Format, 285, 289
HTML, 285, 289, 698
Paßwortschutz, 305
PowerPoint-Format, 285, 289
Publizieren im Internet, 710
Speichern der Textausgabe im Textformat, 316
Textausgabe, 316
Textformat, 285, 290, 698
Word-Format, 285, 289
Speichern von Ansichten von Pivot-Tabellen, 326
Speichern von Dateien, 69, 74, 90
Datenbankdatei-Abfragen, 48
Datendateien, 74, 90
SPSS-Datendateien, 69
Speichern von Diagrammen, 285, 292–295, 297, 678, 698
BMP-Dateien, 285, 292–293, 698
EPS-Dateien, 285, 292, 295, 698
JPEG-Dateien, 285, 292, 698
Metadateien, 285, 292, 698
PICT-Dateien, 285, 292–293, 698
PNG-Dateien, 294
PostScript-Dateien, 295
Speichern von interaktiven Diagrammen mit Daten, 678
TIFF-Dateien, 294
WMF-Dateien, 285, 292, 297
WMF-Format, 698
Speicherplatz, 78, 81
temporär, 78, 81
Speicherzuweisung
in der Two-Step-Clusteranalyse, 537
Spezielle Währungsformate, 101, 683
Split-Half-Reliabilität
in der Reliabilitätsanalyse, 621, 623
Sprache
Ändern der Ausgabesprache, 666
SPSS
grundlegende Schritte, 14
SPSS-Datendateiformat
Ausgabe an eine Datendatei weiterleiten, 757, 763
SPSS-Produktionsmodus, 666, 668–669, 695, 698, 700, 702–704, 707, 709
Angeben eines Remote-Servers, 703
Ausführen mehrerer Produktionsjobs, 707
Ausgabedateien, 695
Befehlszeilenschalter, 707
Ersetzen von Werten in Syntaxdateien, 700
Exportieren von Ausgaben, 698
Exportieren von Diagrammen, 695, 698
Festlegen des Formats über die Befehlssyntax, 706
Festlegen von Formaten für Pivot-Tabellen, 704
Makro-Eingabeaufforderungen, 702
Optionen, 703
Planen von Produktionsjobs, 707
Publizieren im Internet, 709
Publizieren von Ausgaben, 709

Syntaxregeln, 695
Verwenden von Befehlssyntax aus dem Log, 668–669
Verwenden von Befehlssyntax aus der Journaldatei, 666
SPSSTMPDIR (Umgebungsvariable), 666
Stack (Registerkarte), 742
Skript-Fenster, 742
Standardabweichung
in Bericht in Spalten, 615
in Bericht in Zeilen, 608
in der Explorativen Datenanalyse, 381
in Deskriptive Statistiken, 373
in GLM - Univariat, 458
in Häufigkeiten, 366
in Mittelwerte, 410
in OLAP-Würfel, 416
in Verhältnisstatistiken, 637
in Zusammenfassen, 403
Standardfehler
in der Explorativen Datenanalyse, 381
in Deskriptive Statistiken, 373
in GLM, 456, 458
in Häufigkeiten, 366
in ROC-Kurve, 656
Standardfehler der Kurtosis
in Mittelwerte, 410
in OLAP-Würfel, 416
in Zusammenfassen, 403
Standardfehler der Schiefe
in Mittelwerte, 410
in OLAP-Würfel, 416
in Zusammenfassen, 403
Standardfehler des Mittelwerts
in Mittelwerte, 410
in OLAP-Würfel, 416
in Zusammenfassen, 403
Standardisierte Residuen
in GLM, 456
in Lineare Regression, 489
Standardisierte Werte
in Deskriptive Statistiken, 371
Standardisierung
in der Two-Step-Clusteranalyse, 537
Starter-Skripts, 718
Startwert für Zufallszahlen, 170
Stata-Dateien, 29
einlesen, 25
öffnen, 29
speichern, 74
Statistik-Assistent, 15
Statusleiste, 9
Ausblenden, 10
Einblenden, 10
Stengel-Blatt-Diagramme
in der Explorativen Datenanalyse, 382
Stichprobe
Zufallsstichprobe, 241
Streng paralleles Modell
in der Reliabilitätsanalyse, 621, 623
Streß
in Multidimensionale Skalierung, 627
Streudiagramme
in Lineare Regression, 487
Streuungskoeffizient (COD)
in Verhältnisstatistiken, 637
Streuungsmaße
in der Explorativen Datenanalyse, 381
in Deskriptive Statistiken, 373
in Häufigkeiten, 366
in Verhältnisstatistiken, 637
String-Format, 101

String-Variablen, 106, 111
Berechnen von neuen String-Variablen, 168
Eingeben von Daten, 111
fehlende Werte, 106
in Dialogfeldern, 10
lange String-Variablen in früheren Versionen aufteilen, 74
umkodieren in fortlaufende Ganzzahlen, 184
Student-Newman-Keuls-Prozedur
in einfaktorieller ANOVA, 437
in GLM, 453
Student-*T*-Test, 421
Studentisierte Residuen
in Lineare Regression, 489
Studie mit Fallkontrolle
T-Test bei gepaarten Stichproben, 425
Studie mit zugeordneten Paaren
in T-Test bei gepaarten Stichproben, 425
Summe
in Deskriptive Statistiken, 373
in Häufigkeiten, 366
in Mittelwerte, 410
in OLAP-Würfel, 416
in Zusammenfassen, 403
Symbole, 693
Bearbeiten von Bitmap-Symbolen für Symbolleisten, 693
in Dialogfeldern, 13
Symbolleisten, 689–693
Anpassen, 689, 691
Anzeigen in verschiedenen Fenstern, 690
Bearbeiten von Bitmap-Symbolen, 693
ein- und ausblenden, 689
erstellen, 689, 691
Erstellen von neuen Symbolen, 692
Syntax, 353, 361, 666, 668–669, 691, 695, 743, 745
ausführen, 361
Ausführen von Befehlssyntax mit Symbolleisten-Schaltflächen, 691
Ausgabe-Log, 357
einfügen, 356
Einfügen in Skripts, 745
Journaldatei, 359, 362, 666
Log, 668–669
mit Skripts, 743, 745
Regeln für den SPSS-Produktionsmodus, 695
Syntaxregeln, 353
Zugreifen auf *SPSS Command Syntax Reference*, 17
SYSTAT-Dateien, 25
öffnen, 25

T-Test
in GLM - Univariat, 458
in T-Test bei einer Stichprobe, 428
in T-Test bei gepaarten Stichproben, 425
in T-Test bei unabhängigen Stichproben, 421
T-Test bei einer Stichprobe, 428
Fehlende Werte, 430
Konfidenzintervalle, 430
Optionen, 430
zusätzliche Funktionen beim Befehl, 431
T-Test bei gepaarten Stichproben, 425
Auswählen von gepaarten Variablen, 425
Fehlende Werte, 428
Optionen, 428
T-Test bei unabhängigen Stichproben, 421
fehlende Werte, 425
Gruppen definieren, 424
Gruppenvariablen, 424

Konfidenzintervalle, 425
Optionen, 425
String-Variablen, 424
T-Test bei zwei Stichproben
in T-Test bei unabhängigen Stichproben, 421
T2 nach Tamhane
in einfaktorieller ANOVA, 437
in GLM, 453
T3 nach Dunnett
in einfaktorieller ANOVA, 437
in GLM, 453
T4253H-Glättung, 210
Tabellen, 350
Festlegen von Tabellenumbrüchen, 350
Tabellenkalkulationsdateien, 25, 27–28, 78
Einlesen von Bereichen, 27
Einlesen von Variablennamen, 27
öffnen, 28
Schreiben von Variablennamen, 78
Tabellenumbrüche, 350
Tabellenuntertypen, 753
im Vergleich mit Beschriftungen, 755
Tabellenvorlagen, 330–331
erstellen, 331
zuweisen, 331
Tabulator-getrennte Dateien, 25, 27, 50, 74, 78
Einlesen von Variablennamen, 27
öffnen, 25
Schreiben von Variablennamen, 78
speichern, 74
Tau-*b*
in Kreuztabellen, 391
Tau-*c*
in Kreuztabellen, 391
Teilmengen von Fällen
auswählen, 238, 240, 242
Zufallsstichprobe, 241
Temporäre aktive Datei, 81
Temporärer Speicherplatz, 78, 81
Test auf Binomialverteilung, 566
Dichotomien, 566
Fehlende Werte, 568
Optionen, 568
Statistiken, 568
zusätzliche Funktionen beim Befehl, 569
Test Extremreaktionen nach Moses
in Tests bei zwei unabhängigen Stichproben, 578
Tests auf Homogenität der Varianzen.
in einfaktorieller ANOVA, 440
in GLM - Univariat, 458
Tests auf Normalverteilung
in der Explorativen Datenanalyse, 382
Tests auf Unabhängigkeit
Chi-Quadrat, 391
Tests bei mehreren unabhängigen Stichproben, 584
Definieren des Bereichs, 587
fehlende Werte, 587
Gruppenvariablen, 587
Optionen, 587
Statistiken, 587
Testtypen, 586
zusätzliche Funktionen beim Befehl, 588
Tests bei mehreren verbundenen Stichproben, 588
Statistiken, 590
Testtypen, 590
zusätzliche Funktionen beim Befehl, 591
Tests bei zwei unabhängigen Stichproben, 576
fehlende Werte, 579
Gruppen definieren, 579
Gruppenvariablen, 579
Optionen, 579
Statistiken, 579

Testtypen, 578
zusätzliche Funktionen beim Befehl, 580
Tests bei zwei verbundenen Stichproben, 580
fehlende Werte, 583
Optionen, 583
Statistiken, 583
Testtypen, 582
zusätzliche Funktionen beim Befehl, 584
Text, 50, 279–280, 285, 290, 307, 316, 348, 698
Datendateien, 50
einfügen im Viewer, 279
Erstellen von Textausgaben, 307
Exportieren von Ausgaben als Text, 285, 290, 698
Exportieren von Textausgabe im Textformat, 316
Hinzufügen von Textdateien im Viewer, 280
in Zellen, 348
Text-Viewer, 307–309, 314–316, 666, 669
Ändern der Schriftarten, 314
Anzeigeoptionen, 669
Ausgabeformat, 309
Drucken, 314–315
Festlegen der Standardanzeige für Ausgaben, 666
Festlegen des Standard-Ausgabetyps, 308
Rahmenlinien, 309
Spaltenrahmen, 309
Speichern der Ausgabe, 316
Zeilenrahmen, 309
Zellenrahmen, 309
TIFF-Dateien, 294
Exportieren von Diagrammen, 285, 292, 294, 698
Tips für Skripts, 711, 716, 718, 725–726, 728, 730–732, 735, 740
benutzerdefinierte Dialogfelder, 735
Eigenschaften und Methoden, 730
Fehlersuche, 740
Funktionsweise von Skripts, 725
Hinzufügen einer Beschreibung, 735
Objektkatalog, 731
Prozeduren, 732
Skript-Fenster, 716
Starter-Skripts, 718
Variablendeklarationen, 726
Zuweisen von Automatisierungsobjekten, 728
Titel, 279
einfügen im Viewer, 279
in Diagrammen, 647
in OLAP-Würfel, 420
Toleranz
in Lineare Regression, 493
Transformationsmatrix
in der Faktorenanalyse, 517
Transponieren von Variablen und Fällen, 223
Transponieren von Zeilen und Spalten, 319
Trendbereinigte Normalverteilungsdiagramme
in der Explorativen Datenanalyse, 382
Tukey-*B*-Test
in einfaktorieller ANOVA, 437
in GLM, 453
Tukey-Biweight-Schätzer
in der Explorativen Datenanalyse, 381
Tukey-Schätzungen, 182
Tukeys Additivitätstest
in der Reliabilitätsanalyse, 621, 623
Two-Step-Clusteranalyse, 533
Diagramme, 540
in Arbeitsdatei speichern, 541
in externer Datei speichern, 541
Optionen, 537
Statistiken, 541

Übersicht, 781
Umbenennen von Daten-Sets, 124
Umgebungsvariablen, 666
SPSSTMPDIR, 666
Umkodieren von Werten, 152, 174–175, 177–178, 184
Umordnen von Zeilen und Spalten, 320
Umstrukturieren von Daten, 245–246, 249–252, 255, 257–258, 260–261, 263, 265, 267, 269
Auswählen von Daten für "Fälle zu Variablen", 263
Auswählen von Daten für "Variablen zu Fälle", 252
Beispiel für "Fälle zu Variablen", 250
Beispiel für "Variablen zu Fälle", 249
Beispiel für einen Index bei "Variablen zu Fälle", 257
Beispiel für zwei Indizes bei "Variablen zu Fälle", 257
Erstellen einer einzelnen Indexvariable für "Variablen zu Fälle", 258
Erstellen mehrerer Indexvariablen für "Variablen zu Fälle, 260
Erstellen von Indexvariablen für "Variablen zu Fälle", 255
Optionen für "Variablen zu Fälle", 261, 267
Sortieren von Daten für "Fälle zu Variablen", 265
Übersicht, 245
Umstrukturierungstypen, 246
und gewichtete Daten, 269
Variablengruppen für "Variablen zu Fälle", 251
Unähnlichkeitsmaße nach Lance und Williams, 477
in Distanzen, 477
Ungewichtete kleinste Quadrate
in der Faktorenanalyse, 524
Unsicherheitskoeffizient
in Kreuztabellen, 391
Untertitel
in Diagrammen, 647
Untertypen, 753
im Vergleich mit Beschriftungen, 755

V
in Kreuztabellen, 391
Van der Waerden-Schätzungen, 182
Variablen, 12, 98, 115–116, 246, 657, 659, 666
Anzeigereihenfolge in Dialogfeldern, 666
definieren, 98
Definieren von Variablen-Sets, 659
Einfügen von neuen Variablen, 115
in Dialogfeldern, 10
in Dialogfeldern auswählen, 12
Informationen zu Variablen in Dialogfeldern, 13
Informationen zur Definition, 657
Umbenennen für zusammengefügte Datendateien, 228
umkodieren, 174–175, 177–178, 184
Umstrukturieren in Fälle, 246
Verschieben, 116
Variablen-Sets, 659, 661
definieren, 659
verwenden, 661
Variablenansicht, 97
Variablenattribute, 108–109
kopieren und einfügen, 108–109
Variablenbeschreibung, 657
Variablendeklarationen, 726–727
in Skripts, 726–727
Konventionen für die Benennung, 727
Variablenlabels, 104, 666, 671
Einfügen von Zeilenumbrüchen, 105

im Gliederungsfenster, 671
in Dialogfeldern, 11, 666
in Pivot-Tabellen, 671
in zusammengefügten Datendateien, 228
Variablenlisten, 662
Umsortieren von Listen der Zielvariablen, 662
Variablennamen, 99, 666
durch das OMS erzeugt, 772
in Dialogfeldern, 11, 666
lange Variablennamen in früheren Versionen kürzen, 74
portable Dateien, 74
Regeln, 99
Variablennamen für gemischte Fälle, 99
Zeilenumbruch bei langen Variablennamen in der Ausgabe, 99
Variablenpaare, 246
erstellen, 246
Varianz
in Bericht in Spalten, 615
in Bericht in Zeilen, 608
in der Explorativen Datenanalyse, 381
in Deskriptive Statistiken, 373
in Häufigkeiten, 366
in Mittelwerte, 410
in OLAP-Würfel, 416
in Zusammenfassen, 403
Varianz-Inflationsfaktor
in Lineare Regression, 493
Varianzanalyse
in einfaktorieller ANOVA, 433
in Kurvenanpassung, 497
in Lineare Regression, 493
in Mittelwerte, 410
Variationskoeffizient (COV)
in Verhältnisstatistiken, 637
Varimax-Rotation
in der Faktorenanalyse, 527
Verallgemeinerte kleinste Quadrate
in der Faktorenanalyse, 524
Verarbeitung von aufgeteilten Dateien, 237
Verbergen (Ausschließen) der Ausgabe aus dem Viewer bei OMS, 763
Verbundene Stichproben, 580, 588
Vergleichen von Gruppen
in OLAP-Würfel, 419
Vergleichen von Variablen
in OLAP-Würfel, 419
Verhältnisstatistik, 635
Statistiken, 637
Verschieben von Zeilen und Spalten, 320
Verteilter Modus, 83–86, 88, 90–91, 93, 703
Speichern von Datendateien, 90
SPSS-Produktionsmodus, 703
UNC-Pfade, 93
verfügbare Prozeduren, 93
Zugriff auf Datendateien, 88, 91
Verteilungsmaße
in Deskriptive Statistiken, 373
in Häufigkeiten, 366
Vertikaler Beschriftungstext, 321
Verzeichnis für temporäre Dateien, 666
Festlegen des Speicherorts bei lokalen Analysen, 666
SPSSTMPDIR (Umgebungsvariable), 666
Viewer, 271, 273–276, 278–279, 284, 304–305, 668, 671
Abstand zwischen Ausgabeobjekten, 304
Ändern der Größe von Gliederungsobjekten, 279
Ändern der Schriftart in der Gliederung, 279
Ändern von Gliederungsebenen, 278
Anzeigen von Datenwerten, 671

Anzeigen von Variablenlabels, 671
Anzeigen von Variablennamen, 671
Anzeigen von Wertelabels, 671
Anzeigeoptionen, 668
ausblenden von Ergebnissen, 273
Ausgabetypen bei OMS ausschließen, 763
Einfügen spezieller Objekte, 284
Ergebnisfenster, 271
Erweitern der Gliederung, 278
Gliederung, 276
Gliederungsfenster, 271
Kopieren der Ausgabe, 275
Löschen von Ausgaben, 274
Reduzieren der Gliederung, 278
Speichern von Dokumenten, 305
Verschieben der Ausgabe, 274
Vorhergesagte Werte
Speichern in Kurvenanpassung, 502
Speichern in Lineare Regression, 489
Vorhersageintervalle
Speichern in Kurvenanpassung, 502
Speichern in Lineare Regression, 489
Vorlagen, 108–109, 673
in Diagrammen, 651, 673
Variablendefinition, 108–109
Verwenden einer externen Datendatei als Vorlage, 138
Vorlauffunktion, 210
Vorwärtsselektion
in Lineare Regression, 485
Vorzeichentest
in Tests bei zwei verbundenen Stichproben, 580

Wachstumsmodell
in Kurvenanpassung, 501
Währungsformate, 683
Wald-Wolfowitz-Sequenzen
in Tests bei zwei unabhängigen Stichproben, 578
Waller-Duncan-*T*-Test
in einfaktorieller ANOVA, 437
in GLM, 453
Watch (Registerkarte), 742
Skript-Fenster, 742
Wechselwirkungsterme, 448
Welch-Statistik
in einfaktorieller ANOVA, 440
Werte, 343
Anzeigeformat für Pivot-Tabellen, 343
Wertelabels, 104, 112, 118, 126, 671
Einfügen von Zeilenumbrüchen, 105
im Daten-Editor, 118
im Gliederungsfenster, 671
in Pivot-Tabellen, 671
in zusammengefügten Datendateien, 228
Kopieren, 132
mehreren Variablen zuweisen, 132
Verwenden bei der Dateneingabe, 112
Wichtigkeitsdiagramm
in der Two-Step-Clusteranalyse, 540
Wichtigkeitsdiagramme für Variablen
in der Two-Step-Clusteranalyse, 540
Wiederholte Kontraste
in GLM, 450
Wilcoxon-Test
in Tests bei zwei verbundenen Stichproben, 580
Wilks-Lambda
in der Diskriminanzanalyse, 511
Wissenschaftliche Notation, 101, 343, 666
in Pivot-Tabellen, 343
unterdrücken in der Ausgabe, 666

WMF-Dateien, 285, 292, 297, 698
  Exportieren von Diagrammen, 285, 292, 297, 698
Word-Format
  Exportieren von Ausgaben, 285, 289

XML
  Ausgabe an XML weiterleiten, 757
  OXML-Ausgabe aus OMS, 779
  Speichern von Ausgaben als XML, 747
  Tabellenstruktur in OXML, 774
XSLT
  Verwenden mit OXML, 779

Z-Werte
  beim Bilden der Rangfolge, 182
  in Deskriptive Statistiken, 371
  Speichern als Variablen, 371
Zählen von Werten, 171
Zeilen, 348
  Auswählen in Pivot-Tabellen, 348
Zeilenprozente
  in Kreuztabellen, 395
Zeilenumbrüche, 333
  Beeinflussen der Spaltenbreite bei Texten mit Zeilenumbrüchen, 333
  Variablen- und Wertelabels, 105
Zeitreihenanalyse
  Prognose, 502
  Vorhersagen von Fällen, 502
Zeitreihendaten
  Datentransformationen, 206
  Definieren von Datumsvariablen, 206
  Ersetzen fehlender Werte, 211
  Erstellen neuer Zeitreihenvariablen, 208
  Transformationsfunktionen, 210
Zellen in Pivot-Tabellen, 329, 335, 339–340, 342, 344–346, 348
  Ausblenden, 327
  Ausrichtung, 344
  Bearbeiten von Texten, 348
  Breiten, 340
  Einblenden, 327
  Formate, 335
  Gestaltung, 346
  Ränder, 345
  Schattierung, 346
  Schriftarten, 339
  Werteformate, 342
Zellenrahmen, 309
  Text-Viewer, 309
Zentrieren der Ausgabe, 275, 668
Zufallsstichprobe, 40
  auswählen, 241
  Datenbankdateien, 40
  Startwert für Zufallszahlen, 170
Zurücksetzen (Schaltfläche), 11
Zusammenfassen, 399
  Optionen, 402
  Statistiken, 403
Zusammenfügen von Datendateien
  Dateien mit unterschiedlichen Fällen, 224
  Dateien mit unterschiedlichen Variablen, 229
  Informationen aus dem Datenlexikon, 228
  Umbenennen von Variablen, 228
Zusammengesetztes Wachstumsmodell
  in Kurvenanpassung, 501
Zusammenhang linear-mit-linear
  in Kreuztabellen, 391
Zwischenergebnisse
  in Berichte in Spalten, 617